

CORRECTED  
VERSION\*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

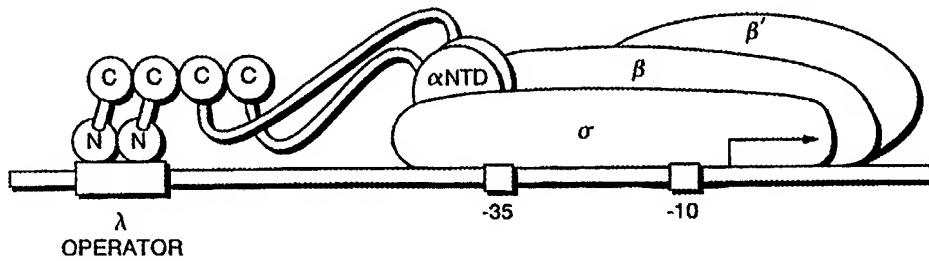


AG  
65

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : C12N 15/10, C12Q 1/68, C12N 1/21	A1	(11) International Publication Number: <b>WO 98/07845</b> (43) International Publication Date: 26 February 1998 (26.02.98)
(21) International Application Number: PCT/US97/14860		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 22 August 1997 (22.08.97)		
(30) Priority Data: 60/024,484 23 August 1996 (23.08.96) US		
(71) Applicant (for all designated States except US): PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).		
(72) Inventors; and		Published
(75) Inventors/Applicants (for US only): DOVE, Simon [US/US]; 205 Walden Street #3N, Cambridge, MA 02140 (US). JOUNG, J., Keith [US/US]; Harvard University, Winthrop House, 32 Mill Street, Cambridge, MA 02138 (US). HOCHSCHILD, Ann [US/US]; 36 Wavery Street, Brookline, MA 02146 (US).		With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot LLP, One Post Office Square, Boston, MA 02109 (US).		

(54) Title: INTERACTION TRAP ASSAY, REAGENTS AND USES THEREOF



(57) Abstract

The present invention makes available an interaction trap system (hereinafter "ITS") which is derived using recombinantly engineered prokaryotic cells.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Iceland	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Interaction Trap Assay, Reagents and Uses Thereof*Funding*

Work described herein was supported by National Institutes of Health Grant. The  
5 United States Government has certain rights in the invention.

*Background of the Invention*

Specific protein-protein interactions are fundamental to most cellular functions. Polypeptide interactions are involved in, *inter alia*, formation of functional transcription  
10 complexes, signal transduction pathways, cytoskeletal organization (e.g., microtubule polymerization), polypeptide hormone receptor-ligand binding, organization of multi-subunit enzyme complexes, and the like.

Investigation of protein-protein interactions under physiological conditions has been problematic. Considerable effort has been made to identify proteins that bind to proteins of  
15 interest. Typically, these interactions have been detected by using co-precipitation experiments in which an antibody to a known protein is mixed with a cell extract and used to precipitate the known protein and any proteins which are stably associated with it. This method has several disadvantages, such as: (1) it only detects proteins which are associated in cell extract conditions rather than under physiological, intracellular conditions, (2) it only  
20 detects proteins which bind to the known protein with sufficient strength and stability for efficient co-immunoprecipitation, (3) may not be able to detect oligomers of the target, and (4) it fails to detect associated proteins which are displaced from the known protein upon antibody binding. Additionally, the precipitation techniques at best provide a molecular weight as the sole identifying characteristic. For these reasons and others, improved  
25 methods for identifying proteins which interact with a known protein have been developed.

One approach has been to use a so-called interaction trap system (also referred to as the "two-hybrid assay") based in yeast to identify polypeptide sequences which bind to a predetermined polypeptide sequence present in a fusion protein (Fields and Song (1989) *Nature* 340:245). This approach identifies protein-protein interactions *in vivo* through  
30 reconstitution of a eukaryotic transcriptional activator.

- 2 -

The interaction trap systems of the prior art are based on the finding that most eukaryotic transcription activators are modular. Brent and Ptashne showed that the activation domain of yeast GAL4, a yeast transcription factor, could be fused to the DNA binding domain of *E. coli* LexA to create a functional transcription activator in yeast (Brent et al. (1985) *Cell* 43:729-736). There is evidence that transcription can be activated through the use of two functional domains of a transcription factor: a domain that recognizes and binds to a specific site on the DNA and a domain that is necessary for activation. The transcriptional activation domain is thought to function by contacting other proteins involved in transcription. The DNA-binding domain appears to function to position the transcriptional activation domain on the target gene that is to be transcribed. These and similar experiments (Keegan et al. (1986) *Science* 231:699-704) formally define activation domains as portions of proteins that activate transcription when brought to DNA by DNA binding domains. Moreover, it was discovered that the DNA binding domain does not have to be physically on the same polypeptide as the activation domain, so long as the two separate polypeptides interact with one another. (Ma et al. (1988) *Cell* 55:443-446).

Fields and his coworkers made the seminal suggestion that protein interactions could be detected if two potentially interacting proteins were expressed as chimeras. In their suggestion, they devised a method based on the properties of the yeast Gal4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation. Polynucleotides encoding two hybrid proteins, one consisting of the yeast Gal4 DNA-binding domain fused to a polypeptide sequence of a known protein and the other consisting of the Gal4 activation domain fused to a polypeptide sequence of a second protein, are constructed and introduced into a yeast host cell. Intermolecular binding between the two fusion proteins reconstitutes the Gal4 DNA-binding domain with the Gal4 activation domain, which leads to the transcriptional activation of a reporter gene (e.g., lacZ, HIS3) which is operably linked to a Gal4 binding site.

All yeast-based interaction trap systems in the art share common elements (Chien et al. (1991) *PNAS* 88:9578-82; Durfee et al. (1993) *Genes & Development* 7:555-69; Gyuris et al. (1993) *Cell* 75:791-803; and Vojtek et al. (1993) *Cell* 74:205-14). All use (1) a plasmid that directs the synthesis of a "bait": a known protein which is brought to DNA by being fused to a DNA binding domain, (2) one or more reporter genes ("reporters") with

- 3 -

upstream binding sites for the bait, and (3) a plasmid that directs the synthesis of proteins fused to activation domains and other useful moieties ("prey"). All current systems direct the synthesis of proteins that carry the activation domain at the amino terminus of the fusion, facilitating the expression of open reading frames encoded by, for example, cDNAs.

5 The prior art systems differ in their specifics. These details are typically relevant to their successful use. Baits differ in their DNA binding domains. For example, systems use baits that contain native *E. coli* LexA repressor protein (Durfee et al. (1993) *Genes & Development* 7:555-69; Gyuris et al. (1993) *Cell* 75:791- 803). LexA binds tightly to appropriate operators (Golemis et al. (1992) *Mol. Cell. Biol.* 12:3006-3014; Ebina et al.

10 (1983) *J. Biol. Chem.* 258:13258-13261), and carries a dimerization domain at its C terminus (Brent R. (1982) *Biochimie* 64:565-569; Little J et al. (1982) *Cell* 29:11-22; and Thliveris et al. (1991) *Biochime* 73:449-455). In yeast, LexA and most LexA derivatives enter the nucleus, but are not necessarily nuclear localized. Others use baits that contain a portion of the yeast GAL4 protein (Chien et al. (1991) *PNAS* 88:9578-82; Durfee et al.

15 (1993) *Genes & Development* 7:555-69; and Harper et al. (1993) *Cell* 75:805-16). This portion, encoded by residues 1-147, is sufficient to bind tightly to appropriate DNA binding sites, localize fused proteins to the nucleus, and direct dimerization; it also contains a domain that weakly activates transcription from mammalian cell extracts *in vitro*, and it is thus conceivable that this domain may increase transcription resulting from weakly

20 interacting proteins.

Reporter genes differ in the phenotypes they confer. The products of some reporter genes (e.g., HIS3, LEU2) allow cells expressing them to be selected by growth on appropriate media, while the products of others (e.g. lacZ) allow cells expressing them to be visually screened. Reporters also differ in the number and affinity of upstream binding sites (e.g., *lexA* operators) for the bait, and in the position of these sites relative to the transcription startpoint (Gyuris et al., *supra*). Finally, they differ in the number of molecules of the reporter gene product necessary to score the phenotype. These differences affect the strength of the protein interactions the reporters can detect .

Preys differ in the activation domains they carry, and in whether they contain other useful moieties such as nuclear localization sequences and epitope tags. Some activation domains are stronger than others. Although strong activation domains should allow

**SUBSTITUTE SHEET (RULE 26)**

- 4 -

detection of weaker interactions, their expression can also harm the cell due to poorly understood transcriptional effects, either by titration of cofactors necessary for transcription of other genes ("squelching") (Gill et al. (1988) *Nature* 334:721-724) or by toxic effects that result when strong activation domains are brought to DNA (Berger et al. (1990) *Cell* 61:1199-208). Thus, it is possible that strong activation domains may prevent detection of some interactions. Activation tagged proteins also differ in whether they are expressed constitutively, or conditionally. Conditional expression allows the transcription phenotypes obtained in selections (or "hunts") for interactors to be ascribed to the synthesis of the tagged protein, thus reducing the number of false positive cells that grow because their 10 reporters are aberrantly transcribed.

Although most two hybrid systems use yeast, there are also mammalian variants. In one, interaction of activation tagged VP16 derivatives with a Gal4-derived bait drives expression of reporters that direct the synthesis of Hygromycin B phosphotransferase, Chloramphenicol acetyltransferase, or CD4 cell surface antigen (Fearon et al. (1992) *PNAS* 89:7958-62). In the other, interaction of VP16-tagged derivatives with Gal4-derived baits drives the synthesis of SV40 T antigen, which in turn promotes the replication of the prey plasmid, which carries an SV40 origin (Vasavada et al. (1991) *PNAS* 88:10686-90).

Several industrially significant uses of two hybrid systems have emerged. One use is to identify new protein targets for pharmaceutical intervention. Typically, the two-hybrid 20 method is used to identify novel polypeptide sequences which interact with a known protein (Silver et al. (1993) *Mol. Biol. Rep.* 17:155; Durfee et al. (1993) *Genes Devel.* 7:555; Yang et al. (1992) *Science* 257:680; Luban et al. (1993) *Cell* 73:1067; Hardy et al. (1992) *Genes Devel.* 6: 801; Bartel et al. (1993) *Biotechniques* 14:920; and Vojtek et al. (1993) *Cell* 74:205). Variations of the two-hybrid method have been used to identify mutations of a 25 known protein that affect its binding to a second known protein (Li B and Fields S (1993) *FASEB J.* 7:957; Lalo et al. (1993) *PNAS* 90:5524; Jackson et al. (1993) *Mol. Cell. Biol.* 13:2899; and Madura et al. (1993) *J. Biol. Chem.* 268:12046). Two-hybrid systems have also been used to identify interacting structural domains of two known proteins (Bardwell et al. (1993) *Med. Microbiol.* 8:1177; Chakraborty et al. (1992) *J. Biol. Chem.* 267:17498; 30 Staudinger et al. (1993) *J. Biol. Chem.* 268:4608; and Milne et al. (1993) *Genes Devel.* 7:1755) or domains responsible for oligomerization of a single protein (Iwabuchi et al.

- 5 -

(1993) *Oncogene* 8:1693; Bogerd et al. (1993) *J. Virol.* 67:5030). Variations of two-hybrid systems have been used to study the *in vivo* activity of a proteolytic enzyme (Dasmahapatra et al. (1992) *PNAS* 89:4159).

5

#### *Summary of the Invention*

The present invention provides methods and reagents for practicing various forms of an interaction trap assay using prokaryotic host cells, e.g., bacterial cells.

For example, one aspect of the present invention relates to a method for detecting 10 interaction between a first test polypeptide and a second test polypeptide. The method comprises a step of providing an interaction trap system including a prokaryotic host cell which contains a reporter gene operably linked to a transcriptional regulatory sequence which includes a binding site ("DBD recognition element") for a DNA-binding domain. The cell is engineered to include a first chimeric gene which encodes a first fusion protein, 15 the first fusion protein including a DNA-binding domain and first test polypeptide. The cell also includes a second chimeric gene which encodes a second fusion protein including an activation tag (such as a polymerase interaction domain [PID]) which activates transcription of the reporter gene when localized to the vicinity of the DBD recognition element. Interaction of the first fusion protein and second fusion protein in the host cell results in 20 measurably greater expression of the reporter gene. Accordingly, the method also includes the steps of measuring expression of the reporter gene, and comparing the level of expression of the reporter gene to a level of expression in a control interaction trap system in which one of both of the first and second test polypeptides are missing from the first and second fusion proteins and resulting fusion proteins do not interact. A statistically 25 significant increase in the level of expression is indicative of an interaction between the first and second test polypeptide portions of the fusion proteins.

Another aspect of the present invention relates to a kit for detecting interaction between a first test polypeptide and a second test polypeptide. The kit can include a first vector for encoding a first fusion protein ("bait fusion protein"), which vector comprises a 30 first gene including (1) transcriptional and translational elements which direct expression in

- 6 -

a prokaryotic host cell, (2) a DNA sequence that encodes a DNA-binding domain and which is functionally associated with the transcriptional and translational elements of the first gene, and (3) a means for inserting a DNA sequence encoding a first test polypeptide into the first vector in such a manner that the first test polypeptide is capable of being expressed

5 in-frame as part of a bait fusion protein containing the DNA binding domain. The kit will also include a second vector for encoding a second fusion protein ("prey fusion protein"), which comprises a second gene including (1) transcriptional and translational elements which direct expression in a prokaryotic host cell, (2) a DNA sequence that encodes a activation tag, such as a polymerase interaction domain (PID), the activation tag DNA

10 sequence being functionally associated with the transcriptional and translational elements of the second gene, and (3) a means for inserting a DNA sequence encoding the second test polypeptide into the second vector in such a manner that the second test polypeptide is capable of being expressed in-frame as part of a prey fusion protein containing the polymerase interaction domain. Additionally, the kit will include a prokaryotic host cell

15 containing a reporter gene having a binding site ("DBD recognition element") for the DNA-binding domain, wherein the reporter gene expresses a detectable protein when a prey fusion protein interacts with a bait fusion protein bound to the DBD recognition element; the host cell being incapable of expressing a protein having the function of (a) the first marker gene, (b) the second marker gene, (c) the DNA-binding domain, and (d) the

20 polymerase interaction domain. Binding of the first test polypeptide and the second test polypeptide in the host cell results in measurably greater expression of the reporter gene than the simultaneous presence of the DNA-binding domain and the polymerase interaction domain in the absence of an interaction between the first test polypeptide and the second test polypeptide.

25 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature.

30 See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989); *DNA Cloning*, Volumes

- 7 -

I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No. 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

#### *Brief Description of the Figures*

15 Figure 1A illustrates that  $\lambda$ cI binds DNA as a dimer, and pairs of dimers bind cooperatively to adjacent operator sites.

Figure 1B illustrates the transcriptional complexes which may formed with a prey fusion protein resulting from replacement of the  $\alpha$ -CTD (C-terminal domain) with the  $\lambda$ cI-CTD. As described in the appended examples, the hybrid  $\alpha$  gene was generated by 20 replacing the gene segment encoding the  $\alpha$ -CTD with a gene segment encoding the  $\lambda$ cI-CTD. A derivative of the *lac* promoter was also created bearing a single  $\lambda$  operator ( $O_R2$ ) in place of the CRP-binding site (centered 62 bps upstream of the transcription startpoint).

25 Figure 2A illustrates the transcriptional complexes which may formed with a prey fusion protein resulting from replacement of the  $\alpha$ -CTD with the GAL11<sup>P</sup> and a bait protein comprised of the  $\lambda$ cI protein having GAL4 fused at its C-terminus.

Figure 2B is a graph indicating the ability of various fusion proteins of GAL11 and GAL11<sup>P</sup> to function in the subject ITS.

Figure 3A depicts the presence of the  $\omega$  subunit in *E. coli* RNA polymerase complexes.

- 8 -

Figure 3B illustrates a covalent system for the  $\omega$  subunit in a  $\lambda$ cI- $\omega$  fusion protein.

Figure 3C is a graph indicating the ability of the  $\lambda$ cI- $\omega$  fusion protein to drive expression of a reporter gene having a  $\lambda$ cI operator.

Figure 3D an ITS using the  $\omega$  subunit in a GAL11<sup>P</sup>- $\omega$  fusion protein.

5 Figure 3E is a graph showing that co-expression of the GAL11<sup>P</sup>- $\omega$  fusion protein with a  $\lambda$ cI-GAL4 fusion protein can activate the expression of a reporter gene under the transcriptional control of a  $\lambda$ cI operator.

Figure 4 is a table illustrating the relative level of reporter gene expression with various combinations of prey and bait fusion proteins derived with p53 sequences.

10

#### *Detailed Description of the Invention*

The eukaryotic interaction trap system ("ITS"), originally developed by Fields and Song (*Nature* (1989) 340:245) in yeast, is a powerful *in vivo* assay to detect protein-protein interactions. It has already had a large impact on basic and applied biological research. In 15 industry, it is being used to isolate and characterize new targets for drug development. It permits researchers to isolate small organic molecules, peptides, and nucleic acids that may lead to new drugs. Future applications for genome characterization and for modulation of specific protein-protein interactions are on the horizon. The ramifications of this technology promise to be exciting. In this system, one protein is fused to a DNA binding domain, while 20 the other is fused to a transcriptional activating domain. If the two proteins interact in a yeast cell, a functional transcriptional activator is reconstituted, the activity of which is monitored by the expression of a reporter gene containing a cognate site for the DNA binding domain. A number of different DNA binding domains and activation domain have been successfully used in this system, as well as a variety of different reporter genes. 25 However, the interaction trap assays described in the art have only been generated in eukaryotic cells. There are no examples in the art of an analogous system being generated in prokaryotes.

The present invention makes available an interaction trap system (hereinafter "ITS") which is derived using recombinantly engineered prokaryotic cells. As described in the

- 9 -

appended examples, the prokaryotic ITS derives in part from the unexpected finding that the natural interaction between a transcriptional activator and subunit(s) of an RNA polymerase complex can be replaced by a heterologous protein-protein interaction which is capable of activating transcription. The versatility of the prokaryotic ITS makes it generally suitable for many, if not all of the applications of the eukaryotic ITS. Moreover, the ease of manipulation of the bacterial cells, e.g., in transformation or transfection and culturing, means that even larger polypeptide libraries can be sorted in the prokaryotic ITS.

The prokaryotic interaction trap systems described herein provide advantages over the conventional eukaryotic ITS methods. For example, the use of bacterial host cells to generate an interaction trap system provides a system which is generally easier to manipulate genetically relative to the eukaryotic systems. Furthermore, bacterial host cells are easier to propagate. The shorter doubling times for bacteria will often provide for development of a signal in the ITS in a shorter time period than would be obtained with a eukaryotic ITS. Another advantage which may be realized in the practice of the present invention is that detection of reporter gene expression can, in certain embodiments, be technically easier relative to the eukaryotic system. The expression of a  $\beta$ -galactosidase reporter gene, for example, is more easily detected in bacteria than in yeast.

Yet another benefit which may be realized by the use of the prokaryotic ITS is lower spurious activation relative to, e.g., the ITS fusion proteins employed in yeast. In eukaryotic cells, spurious transcription activation by a bait polypeptide having a high acidic residue content can be problematic. This is not expected to be an impediment for the use of such bait polypeptides in the prokaryotic ITS.

Another benefit in the use of the prokaryotic ITS is that, in contrast to the eukaryotic system, nuclear localization of the bait and prey polypeptides is not a concern in bacterial cells.

Still another advantage of the use of the prokaryotic ITS can be realized where the bait and/or prey polypeptides are derived from eukaryotic sources, such as human. One problem which can occur when using the yeast ITS of the prior art is that mammalian/eukaryotic derived bait or prey may retain sufficient biological activity in yeast cells so as to confound the results of the ITS. The greater evolutionary divergence between

- 10 -

mammals and bacteria reduces the likelihood of a similar problem in the prokaryotic ITS of the present invention.

### *I. Overview*

5 A method and reagents for detecting interactions between two polypeptides is provided in accordance with the present invention. The method generally includes, with some variations, providing a recombinant prokaryotic cell engineered to include a reporter gene construct including (i) a binding site ("DBD recognition element") for a DNA-binding domain operably linked to (ii) at least one reporter gene which expresses a reporter gene 10 product when the gene is transcriptionally activated.

The cell is also engineered to include a first chimeric gene which is capable of being expressed in the host cell. The chimeric gene encodes a fusion protein (a "bait" fusion protein) which comprises (i) a DNA-binding domain that specifically binds the recognition element on the reporter gene in the host cell, and (ii) a "bait" polypeptide, e.g., a test 15 polypeptide for which complex formation is to be tested. The DNA-binding domain and bait polypeptide are preferably from heterologous sources.

A second chimeric gene is also provided in the cell, the second chimeric gene encoding a second hybrid protein (a "prey" fusion protein) comprising an "activation tag", e.g., a polypeptide capable of recruiting an active polymerase complex, fused to a test 20 polypeptide sequence (a "prey" polypeptide) which is to be tested for interaction with the bait polypeptide. In certain embodiments of the prokaryotic ITS, the activation tag can be a polymerase interaction domain of an RNA polymerase subunit. For instance, the polymerase interaction domain ("PID") can include determinants of an RNA polymerase subunit that mediate its interaction with other polymerase subunits, thus enabling the prey 25 fusion protein to be assembled into a functional polymerase enzyme.

In other embodiments, the polymerase interaction domain can be a polypeptide sequence which interacts with, or is covalently bound to, one or more subunits (or a fragment thereof) of an RNA polymerase complex in order to recruit functional polymerases to the DNA sequestered prey protein. Such polypeptide sequences can be 30 derived from, e.g., transcription factors or auxiliary proteins of polymerase complexes or

- 11 -

even from random polypeptide libraries (e.g., not occurring naturally). For instance, the prey fusion protein is derived with an activation domain of a transcriptional activator, rather than with the polymerase interaction domain described above. In those embodiments, the prey fusion protein must function to directly or indirectly recruit the RNA polymerase 5 enzyme to the reporter gene by forming bridging contacts to one or more of the polymerase subunits. In either embodiment, expression of the reporter gene occurs when the activation tag is brought into sufficient proximity to the reporter gene by the prey protein contacting a bait protein whose DNA-binding domain is bound to the recognition element.

In one embodiment, both the first and the second chimeric genes are introduced into 10 the host cell in the form of plasmids.

The bait/prey-mediated interaction, if any, between the first and second fusion proteins in the host cell causes an RNA polymerase complex to be recruited to the transcriptional regulatory sequences of the reporter gene with concomitant transcription of the reporter gene. The method is carried out by introducing the first and second chimeric 15 genes into the host cell, and subjecting that cell to conditions under which the first and second hybrid proteins are expressed in sufficient quantity for expression of the reporter gene to be activated by interaction of the two fusion proteins if that interaction occurs. The formation of a complex between the bait and prey fusion proteins results in a detectable signal produced by the expression of the reporter gene. Accordingly, the formation of a 20 complex between a sample target protein and proteins encoded by a cDNA library, for example, can be detected, and ITS cells isolated, if desired, on the basis of evaluating the level of expression of the reporter gene.

The method of the present invention, as described above, may be practiced using a kit for detecting interaction between a first test protein and a second test protein. The kit 25 typically will include the two vectors for generating the chimeric proteins, a reporter gene construct, and a host cell. The first vector contains a promoter and may include a transcription termination signal functionally associated with the first chimeric gene in order to direct the transcription of the first chimeric gene. The first chimeric gene includes a DNA sequence that encodes a DNA-binding domain and a (unique) restriction site(s) for 30 inserting a DNA sequence encoding a first test polypeptide in such a manner that the first test protein is expressed as part of a hybrid protein with the DNA-binding domain. The first

- 12 -

vector also includes a means for replicating itself in the host cell. Also included on the first vector is, preferably, a first marker gene, the expression of which in the host cell permits selection of cells containing the first marker gene. Exemplary marker genes confer antibiotic resistance. Preferably, the first vector is a plasmid.

5        The second vector is derived for generating the second chimeric protein. The second chimeric gene includes a promoter and other relevant transcription and/or translation sequences to direct expression of the chimeric gene. The second chimeric gene also includes a DNA sequence that encodes an activation tag and a (unique) restriction site(s) to insert a DNA sequence encoding the second test polypeptide into the vector, in such a  
10 manner that the second test protein is capable of being expressed as part of a hybrid protein with the activation tag. The second vector further includes a means for replicating itself in the host cell. The second vector also includes a second marker gene, the expression of which in the host cell permits selection of cells containing the second marker gene.

15       The kit includes a prokaryotic host cell, preferably a strain of *E. coli* or other suitable bacterial strain, which can be engineered to express the bait and prey fusion proteins, and express the reporter gene in a manner dependent on the formation of complexes including the two fusion proteins. The host cell contains the reporter gene having a DNA binding site for the DNA-binding domain of the first hybrid protein. The binding site is positioned so that, upon interaction of the bait and prey fusion proteins, an  
20 RNA polymerase complex is recruited to the promoter sequence of the reporter gene, causing expression of the reporter gene. The host cell, by itself, is preferably incapable of expressing a protein having a function of the first marker gene, the second marker gene, the reporter gene, or the complex of the prey and bait fusion proteins.

25       Accordingly, in using the kit the interaction of the bait and prey components of the two fusion proteins in the host cell causes a measurably greater expression of the reporter gene than when the DNA-binding domain and the polymerase interaction domain are provided alone, e.g., without one or both of the bait or prey polypeptides. The reporter gene may encode an enzyme or other product that can be readily measured. Such measurable activity may include the ability of the cell to grow only when the marker gene is  
30 transcribed, or the presence of detectable enzyme activity only when the marker gene is transcribed.

- 13 -

The cells containing the two hybrid proteins are incubated in/on an appropriate medium and the cells are monitored, and optionally selected, by detecting expression of the reporter gene product. Expression of the reporter gene is an indication that the bait protein and the prey protein have interacted.

5

## *II. Definitions*

Before further description of the invention, certain terms employed in the specification, examples and appended claims are, for convenience, collected here.

The term "prokaryote" is art recognized and refers to a unicellular organism lacking 10 a true nucleus and nuclear membrane, having genetic material composed of a single loop of naked double-stranded DNA. Prokaryotes with the exception of mycoplasmas have a rigid cell wall. In some systems of classification, a division of the kingdom Prokaryotae, Bacteria include all prokaryotic organisms that are not blue-green algae (Cyanophyceae). In other systems, prokaryotic organisms without a true cell wall are considered to be unrelated 15 to the Bacteria and are placed in a separate class--the Mollicutes.

The term "bacteria" is art recognized and refers to certain single-celled microorganisms of about 1 micrometer in diameter; most species have a rigid cell wall. They differ from other organisms (eukaryotes) in lacking a nucleus and membrane-bound organelles and also in much of their biochemistry.

20 As used herein, "recombinant cells" include any cells that have been modified by the introduction of heterologous DNA.

As used herein, the terms "heterologous DNA" or "heterologous nucleic acid" is meant to include DNA that does not occur naturally as part of the genome in which it is present, or DNA which is found in a location or locations in the genome that differs from 25 that in which it occurs in nature, or occurs extra-chromosomally, e.g., as part of a plasmid.

By "protein" or "polypeptide" is meant a sequence of amino acids of any length, constituting all or a part of a naturally-occurring polypeptide or peptide, or constituting a non-naturally-occurring polypeptide or peptide (e.g., a randomly generated peptide sequence or one of an intentionally designed collection of peptide sequences).

- 14 -

By a "DNA binding domain" or "DBD" is meant a polypeptide sequence which is capable of directing specific polypeptide binding to a particular DNA sequence (i.e., to a DBD recognition element). The term "domain" in this context is not intended to be limited to a discrete folding domain. Rather, consideration of a polypeptide as a DBD for use in the 5 bait fusion protein can be made simply by the observation that the polypeptide has a specific DNA binding activity. DNA binding domains, like activation tags, can be derived from proteins ranging from naturally occurring proteins to completely artificial sequences.

The term "activation tag" refers to a polypeptide sequence capable of affecting transcriptional activation, for example assembling or recruiting an active polymerase 10 complex. For instance, in the prokaryotic ITS the activation tag can be a polymerase interaction domain or some other polypeptide sequence which interacts with, or is covalently bound to, one or more subunits (or a fragment thereof) of an RNA polymerase complex. Activation tags can also be sequences which are derived from, e.g., transcription factors or auxiliary proteins of polymerase complexes or even from random polypeptide 15 libraries.

The term "polymerase interaction domain" or "PID" are activation tags which include determinants of an RNA polymerase subunit that mediate its interaction with other polymerase subunits, or a polypeptide sequence which interacts with, or is covalently bound to, one or more subunits (or a fragment thereof) of an RNA polymerase complex.

20 The terms "recombinant protein", "heterologous protein" and "exogenous protein" are used interchangeably throughout the specification and refer to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. That is, the polypeptide is expressed from a 25 heterologous nucleic acid.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter 20 gene" operatively linked to transcriptional regulatory sequences. Transcription of the reporter gene is controlled by these sequences. The activity of at least one or more of these control sequences is directly or indirectly regulated by a transcriptional complex recruited 30 by virtue of interaction between the bait and prey fusion proteins. The transcriptional

- 15 -

regulatory sequences can include a promoter and other regulatory regions that modulate the activity of the promoter, or regulatory sequences that modulate the activity or efficiency of the RNA polymerase that recognizes the promoter. Such sequences are herein collectively referred to as transcriptional regulatory elements or sequences. The reporter gene construct 5 will also include a "DBD recognition element" which is a nucleotide sequence that is specifically bound by the DNA binding domain of the bait fusion protein. The DBD recognition element is located sufficiently proximal to the promoter sequence of the reporter gene so as to cause increased reporter gene expression upon recruitment of an RNA polymerase complex by a bait fusion protein bound at the recognition element.

10 As used herein, a "reporter gene" is a gene whose expression may be assayed; reporter genes may encode any protein that provides a phenotypic marker, for example: a protein that is necessary for cell growth or a toxic protein leading to cell death, e.g., a protein which confers antibiotic resistance or complements an auxotrophic phenotype; a protein detectable by a colorimetric/fluorometric assay leading to the presence or absence of 15 color/fluorescence; or a protein providing a surface antigen for which specific antibodies/ligands are available.

By "operably linked" is meant that a gene and transcriptional regulatory sequence(s) are connected in such a way as to permit expression of the gene in a manner dependent upon factors interacting with the regulatory sequence(s). In the case of the reporter gene, the 20 DBD recognition element will also be operably linked to the reporter gene such that transcription of the reporter gene will be dependent, at least in part, upon bait-prey complexes bound to the recognition element.

By "covalently bonded" it is meant that two domains are joined by covalent bonds, directly or indirectly. That is, the "covalently bonded" proteins or protein moieties may be 25 immediately contiguous or may be separated by stretches of one or more amino acids within the same fusion protein.

By "altering the expression of the reporter gene" is meant a statistically significant increase or decrease in the expression of the reporter gene to the extent required for detection of a change in the assay being employed. It will be appreciated that the degree of

- 16 -

change will vary depending upon the type of reporter gene construct or reporter gene expression assay being employed.

The terms "interactors", "interacting proteins" and "candidate interactors" are used interchangeably herein and refer to a set of proteins which are able to form complexes with 5 one another, preferably non-covalent complexes.

By "test protein" or "test polypeptide" is meant all or a portion of one of a pair of interacting proteins provided as part of the bait or prey fusion proteins.

By "randomly generated" is meant sequences having no predetermined sequence; this is contrasted with "intentionally designed" sequences which have a DNA or protein 10 sequence or motif determined prior to their synthesis.

By "amplification" or "clonal amplification" is meant a process whereby the density of host cells having a given phenotype is increased.

The terms "pool" of polypeptides, "polypeptide library" or "combinatorial polypeptide library" are used interchangeably herein to indicate a variegated ensemble of 15 polypeptide sequences, where the diversity of the library may result from cloning or be generated by mutagenesis. The terms "pool" of genes, "gene library" or "combinatorial gene library" have a similar meaning, indicating a variegated ensemble of nucleic acids.

By "screening" is meant a process whereby a gene library is surveyed to determine whether there exists within this population one or more genes which encode a polypeptide 20 having a particular binding characteristic in the interaction trap assay.

It is further noted that the following description of particular arrangements of test polypeptide sequences in terms of being part of the bait or prey fusion proteins is, in general, arbitrary. As will be apparent from the description, the test polypeptide portions of any given pair of interacting bait and prey fusion proteins may ordinarily be swapped with 25 one another.

Each component of the system is now described in more detail.

### *III. Bait protein constructs*

One of the first steps in the use of the interaction trap system of the present invention is to construct the bait fusion protein. To do this, sequences encoding a protein of interest or a polypeptide library are cloned in-frame to a sequence encoding a DNA binding domain (DBD), e.g., a polypeptide which specifically binds to a defined nucleotide sequence. Those skilled in the art will appreciate from the present disclosure that there are a wide variety of DNA binding domains that can be used to construct the bait fusion protein, including polypeptides derived from naturally occurring DNA binding proteins, as well as polypeptides derived from proteins artificially engineered to interact with specific DNA sequences. Basic requirements for the bait fusion protein include the ability to specifically bind a defined nucleotide sequence, and (preferably) that the bait fusion protein cause little or no transcriptional activation of the reporter gene in the absence of an interacting prey fusion protein. In addition, the bait polypeptide sequence should not affect the ability of the DBD to bind to its cognate sequence in the transcriptional regulatory element of the reporter gene.

In one preferred embodiment, the DBD portion of the bait fusion protein is derived using all, or a DNA binding portion of a transcriptional regulatory protein, e.g., of either a transcriptional activator or transcriptional repressor, which retains the ability to selectively bind to particular nucleotide sequences. The DNA binding domains of the bacteriophage  $\lambda$ cI protein (hereinafter " $\lambda$ cI") and the *E. coli* LexA repressor (hereinafter "LexA") represent preferred DNA binding domains for the bait fusion proteins of the instant interaction trap system. The use of a well-defined system, such as  $\lambda$ cI or LexA, allows knowledge regarding the interaction between a DNA binding domain and its DBD recognition element (i.e., the  $\lambda$ cI or LexA operator) to be exploited for the purpose of optimizing operator occupancy and/or optimizing the geometry of the bound bait protein to effect maximal gene activation. In constructing the bait fusion protein, the DNA binding activity of the fusion protein can be, as appropriate, provided by using all or a portion of the transcriptional regulatory protein. Depending on the sequences of the regulatory protein retained in the bait fusion protein, it may be desirable to mutate certain residues of those retained sequences which may contribute to transcriptional activation or repression in the absence of the prey fusion protein, e.g., in order to reduce prey-independent modulation of reporter gene transcription.

- 18 -

However, any other transcriptionally inert or essentially transcriptionally-inert DNA binding domain may be used to create the bait fusion protein in the instant interaction trap system; such DNA binding domains are well known and include, but are not limited to such motifs as helix-turn-helix motifs (such as found in  $\lambda$ cI), winged helix-turn helix motifs 5 (such as found in certain heat shock transcription factors), and/or zinc fingers/zinc clusters. As merely illustrative, the bait fusion protein can be constructed utilizing the DNA binding portions of the LysR family of transcriptional regulators, e.g., TrpI, HvY, OccR, OxyR, CatR, NahR, MetR, CysB, NodD or SyrM (Schell et al. (1993) *Annu Rev Microbiol* 47:597), or the DNA binding portions of the PhoB/OmpR-related proteins, e.g., PhoB, 10 OmpR, CacC, PhoM, PhoP, ToxR, VirG or SfrA (Makino et al. (1996) *J Mol Biol* 259:15), or the DNA binding portions of histones H1 or H5 (Suzuki et al. (1995) *FEBS Lett* 372:215). Other exemplary DBD's which can be used to generate the bait fusion protein include DNA binding portions of the P22 Arc repressor, MetJ, CENP-B, RapI, Xy1S/Ada/AraC, Bir5 or DtxR.

15 Furthermore, the DNA binding domain need not be obtained from the protein of a prokaryote. For example, polypeptides with DNA binding activity can be derived from proteins of eukaryotic origin, including from yeast. For example, the DBD portion of the bait fusion protein can include polypeptide sequences from such eukaryotic DNA binding proteins as p53, jun, fos, GCN4, or GAL4. Likewise, the DNA binding portion of the bait 20 fusion protein can be generated from viral proteins, such as the papillomavirus E2 protein (c.f., PCT publication WO 96/19566). In yet other embodiments, the DNA binding protein can be generated by combinatorial mutagenic techniques, and represent a DBD not naturally occurring in any organism. A variety of techniques have been described in the art for generating novel DNA binding proteins which can selectively bind to a specific DNA 25 sequence (c.f., U.S. Patent 5,198,346 entitled "*Generation and selection of novel DNA-binding proteins and polypeptides*").

As appropriate, the DNA binding motif used to generate the bait fusion protein can include oligomerization motifs. As known in the art, certain transcriptional regulators dimerize, with dimerization promoting cooperative binding of the two monomers to their 30 cognate recognition elements. For example, where the bait protein includes a LexA DNA binding domain, it can further include a LexA dimerization domain; this optional domain

- 19 -

facilitates efficient LexA dimer formation. Because LexA binds its DNA binding site as a dimer, inclusion of this domain in the bait protein also optimizes the efficiency of operator occupancy (Golemis and Brent, (1992) Mol. Cell Biol. 12:3006). Other oligomerization motifs useful in the present invention will be readily recognized by the those skilled in the art. Exemplary motifs include the tetramerization domain of p53 and the tetramerization domain of BCR-ABL. In addition, the art also provides a variety of techniques for identifying other naturally occurring oligomerization domains, as well as oligomerization domains derived from mutant or otherwise artificial sequences. See, for example, Zeng et al. (1997) *Gene* 185:245.

10 As described below, binding efficiency of the bait fusion protein for the recognition element of the reporter gene can also be fine tuned by the particular sequence of the DBD recognition element, and its proximity to other transcriptional regulatory sequences in the reporter gene construct. Likewise, the binding efficiency and/or specificity of the DBD portion of the bait fusion protein can be altered by mutagenesis.

15 The bait portion of the bait fusion protein may be chosen from any protein of interest and includes proteins of unknown, known, or suspected diagnostic, therapeutic, or pharmacological importance. Exemplary bait proteins include, but are not limited to, oncoproteins (such as myc, particularly the C-terminus of myc, ras, src, fos, and particularly the oligomeric interaction domains of fos), tumor-suppressor proteins (such as p53, Rb, 20 INK4 proteins [p16INK4a, p15INK4b], CIP/KIP proteins [p21CIP1, p27KIP1]) or any other proteins involved in cell-cycle regulation (such as kinases and phosphatases). In other embodiments, the bait polypeptide can be generated using all or a portion of a protein involved in signal transduction, including such motifs as SH2 and SH3 domains, ITAMs, ITIMs, kinase, phospholipase, or phosphatase domains, cytoplasmic tails of receptors and 25 the like. Yet other preferred bait fusion proteins are generated with cytoskeletal proteins or factors involved in transcription or translation, or portions thereof. Still other bait fusion proteins can be generated with viral proteins.

In preferred embodiments, where the bait protein includes a catalytic domain of an enzyme, the fusion protein is derived with a catalytically inactive mutant, most preferably a 30 mutant which binds substrate with about the  $K_m$  of the wild-type enzyme but with a greatly diminished  $K_{cat}$  for the catalyzed reaction with the substrate. For example, mutation of a

- 20 -

residue in the catalytic site of the enzyme can give rise to such catalytically inactive mutants. Particular examples include point mutation of the active site lysine of a kinase, the active site serine of a serine protease or the active site cysteine of a phosphatase. Thus, the binding of the bait polypeptide portion of the fusion protein to a polypeptide substrate 5 presented by a prey fusion protein can be enhanced. In each case, the protein of interest is fused to a DNA binding domain as generally described herein.

The use of recombinant DNA techniques to create a fusion gene, with the translational product being the desired bait fusion protein, is well known in the art. Essentially, the joining of various DNA fragments coding for different polypeptide 10 sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. Alternatively, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. In another 15 method, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, Eds. Ausubel et al. John Wiley & Sons: 1992).

It may be necessary in some instances to introduce an unstructured polypeptide 20 linker region between the DNA binding domain of the fusion protein and the bait polypeptide sequence. Where the bait fusion protein also includes oligomerization sequences, it may be preferable to situate the linker between the oligomerization sequences and the bait polypeptide. The linker can facilitate enhanced flexibility of the fusion protein allowing the DBD to freely interact with a responsive element, and, if present, the 25 oligomerization sequences to make inter-protein contacts. The linker can also reduce steric hindrance between the two fragments, and allow appropriate interaction of the bait polypeptide portion with a prey polypeptide component of the interaction trap system. The linker can also facilitate the appropriate folding of each fragment to occur. The linker can be of natural origin, such as a sequence determined to exist in random coil between two 30 domains of a protein. An exemplary linker sequence is the linker found between the C-terminal and N-terminal domains of the RNA polymerase  $\alpha$  subunit. Other examples of

- 21 -

naturally occurring linkers include linkers found in the  $\lambda$ cI and LexA proteins. Alternatively, the linker can be of synthetic origin. For instance, the sequence (Gly<sub>4</sub>Ser)<sub>3</sub> can be used as a synthetic unstructured linker. Linkers of this type are described in Huston et al. (1988) PNAS 85:4879; and U.S. Patent No. 5,091,513, both incorporated by reference 5 herein. Another exemplary embodiment includes a poly alanine sequence, e.g., (Ala)<sub>3</sub>.

As set out above, the bait fusion protein should have little to no transcriptional activation ability by itself. In a preferred embodiment, a repression assay is carried out as a control to confirm that lack of transcriptional activation by the bait fusion protein is not simply because the fusion protein is mis-folded, or is sequestered in occlusion bodies. In 10 one embodiment, the repression assay tests the ability of the fusion protein to competitively block transcription of a reporter gene construct containing a DBD recognition element. For example, a bait fusion protein including a DBD from PhoB can be validated, in part, by observing the ability of the fusion protein to inhibit, in the presence of wild-type PhoB, expression of a reporter gene operably linked to a *pho* box sequence. Where the bait fusion 15 protein includes the DNA binding domain of  $\lambda$ cI, the ability of the fusion protein to bind to a  $\lambda$  operator sequence (e.g., which could serve as the DBD recognition element) can be validated by its ability to confer on an *E. coli* strain immunity to infection by  $\lambda$  phage.

#### *IV. Prey protein constructs*

20 In preferred embodiments, the prey fusion protein comprises: (1) a target polypeptide sequence, capable of forming an intermolecular association with the bait polypeptide which is to be tested for such binding activity, and (2) an activation tag such as a PID. As described herein, the activation tag can be, for example, all or a portion of an RNA polymerase subunit, such as the polymerase interaction domain of the N-terminal 25 domain ( $\alpha$ -NTD) of the RNA polymerase  $\alpha$  subunit. As described above, protein-protein contact between the bait and prey fusion proteins (via the interacting bait and prey polypeptide portions of those proteins) links the DNA-binding domain of the bait fusion protein with the polymerase interaction domain of the prey fusion protein, generating a protein complex capable of directly recruiting a functional RNA polymerase enzyme to 30 DNA sequences proximate to the DNA bound bait protein, i.e., to the reporter gene.

- 22 -

DNA dependent RNA polymerase in *E. coli* and other bacteria consists of an enzymatic core composed of subunits  $\alpha$ ,  $\beta$ , and  $\beta'$  in the stoichiometry  $\alpha_2\beta\beta'$ , and one of several alternative  $\sigma$  factors responsible for specific promoter recognition. In one embodiment, the prey fusion protein includes a sufficient portion of the amino-terminal 5 domain of the  $\alpha$  subunit to permit assembly of transcriptionally active RNA polymerase complexes which include the prey fusion protein. The  $\alpha$  subunit, which initiates the assembly of RNA polymerase by forming a dimer, has two independently folded domains (Ebright et al. (1995) *Curr Opin Genet Dev* 5:197). The larger amino-terminal domain ( $\alpha$ -NTD) mediates dimerization and the subsequent assembly of the polymerase complex. The 10 prey polypeptide can be fused in frame to the  $\alpha$ -NTD (see appended examples) or a fragment thereof which retains the ability to assemble a functional RNA polymerase complex.

To further illustrate the ability of the  $\alpha$  subunit to be utilized in the subject ITS, the coding sequence for  $\alpha$ -NTD was fused to the coding sequence for the yeast protein 15 GAL11<sup>P</sup>, a mutant form of GAL11. See Figure 2A and Himmelfarb et al. (1990) *Cell* 63:1299-309. The "P" mutation confers upon GAL11, a component of the RNA polymerase II holoenzyme in yeast, the ability to interact with a portion of the dimerization region of GAL4. We also constructed a fusion protein comprised of the  $\lambda$ cI protein having 20 GAL4 fused at its C-terminus. As demonstrated in Figure 2B, the co-expression of both fusion proteins can activate the expression of a reporter gene under the transcriptional control of a  $\lambda$ cI operator. Substitution of the wildtype GAL11 sequence for the GAL11<sup>P</sup> sequence result in loss of transcriptional activity of the co-expressed fusion proteins.

Figure 4 similarly illustrates the use of the  $\alpha$ -NTD. In that embodiment, p53 was fused to both  $\alpha$ -NTD and to the DBD of  $\lambda$ cI. The p53 protein includes, in its carboxy 25 terminus, an oligomerization domain which mediates formation of p53 homodimers and heterodimers. As demonstrated in Figure 4, the co-expression of both fusion proteins can activate the expression of a reporter gene under the transcriptional control of a  $\lambda$ cI operator, presumably by p53-mediated oligomerization (e.g., dimerization and/or tetramerization). Expression of only the p53/ $\lambda$ cI, e.g., in the presence of the wildtype  $\alpha$  subunit, did not 30 activate expression of the reporter gene above basal levels.

- 23 -

The present invention also contemplates the use of polymerase interaction domains containing portions of other RNA polymerase subunits or portions of molecules which associate with an RNA polymerase subunit or subunits. Contemporary models of the polymerase complex predict a substantial degree of intramolecular motion within the transcription complex. Movement of parts of the enzyme complex relative to each other is believed to be realized by structurally independent domains, such as the N-terminal and C-terminal domains of the  $\alpha$  subunit described above. Accordingly, it is possible that the paradigm of transcriptional activation realized with fusion proteins incorporating only a portion of the  $\alpha$  subunit is also applicable to fusion proteins generated with portions of other polymerase subunits, preferably subunits which are an integral part of or tightly associated with the polymerase complex, e.g., such as the  $\beta$ ,  $\beta'$ ,  $\omega$  and/or  $\sigma$  subunits. The use of portions of such other subunits to generate a prey fusion protein are, like the  $\alpha$ -NTD example above, expected to provide fusion proteins which retain the ability to form active polymerase complexes. For example, Severinov et al. (1995) PNAS 92:4591 describes the ability of fragments of the  $\beta$  subunit (encoded by the *E. coli* *rpoB* gene) to reconstitute a functional polymerase enzyme. It is noted that it may be a formal requirement of embodiments utilizing prey fusion proteins including PIDS of the  $\beta$ ,  $\beta'$  or  $\sigma$  subunits that other fragments of the subunit be provided, e.g., co-expressed, in the host cell.

To further illustrate such equivalents, it is noted that highly purified *E. coli* RNA polymerase contains a small subunit termed omega ( $\omega$ ). See Figure 3A. This subunit consists of 91 amino acids with a molecular weight of 10,105. Its cloning has been previously reported (Gentry et al. (1986) *Gene* 48:33-40). We fused the  $\omega$  coding sequence in frame to the C-terminus of  $\lambda$ cI. See Figure 3B. In bacterial strains lacking wildtype  $\omega$ , the  $\lambda$ cI- $\omega$  fusion protein was able to drive expression of a  $\beta$ -gal reporter gene having a  $\lambda$ cI operator. Figure 3C illustrates that  $\lambda$ cI itself was unable efficiently induce expression of the reporter gene. Moreover, wildtype  $\omega$  can effectively compete for binding to the holoenzyme complex, and can inhibit the ability of  $\lambda$ cI- $\omega$  to induce expression of the reporter gene.

To demonstrate the ability of the  $\omega$  subunit to be utilized in the subject ITS, the coding sequence for  $\omega$  was fused to the coding sequence for GAL11<sup>P</sup>. See Figure 3D. We also constructed a fusion protein comprised of the  $\lambda$ cI protein having GAL4 fused at its C-terminus. As demonstrated in Figure 3E, the co-expression of both fusion proteins can

- 24 -

activate the expression of a reporter gene under the transcriptional control of a  $\lambda$ cI operator. Substitution of the wildtype GAL11 sequence for the GAL11<sup>P</sup> sequence result in loss of transcriptional activity of the co-expressed fusion proteins.

Additionally, given the general conservation of the polymerase subunits amongst 5 bacteria, the present invention also specifically contemplates prey fusion proteins derived with polymerase interaction domains of RNA polymerase subunits from other bacteria, e.g., *Staphylococcus aureus* (Deora et al. (1995) Biochem Biophys Res Commun 208:610), *Bacillus subtilis*, etc.

In an alternative embodiment, instead of a polymerase interaction domain, the prey 10 fusion protein can include an activation domain of a transcriptional activator protein. The bait fusion protein, by forming DNA bound complexes with the prey fusion protein, can indirectly recruit RNA polymerase complexes to the promoter sequences of the reporter 15 gene, thus activating transcription of the reporter gene. To illustrate, the activation domain can be derived from such transcription factors as PhoB or OmpR. The critical consideration in the choice of the activation domain is its ability to interact with RNA polymerase 20 subunits or complexes in the host cell in such a way as to be able to activate transcription of the reporter gene.

The prey fusion proteins can differ in the polymerase interaction domains or target 25 surfaces they include, and in whether they contain other useful moieties such as epitope tags, oligomerization domain, etc. There are also a wide variety of prey polypeptides which can be selected to generate the fusion protein. The prey polypeptide can be derived from all or a portion of a known protein or a mutant thereof, all or a portion of an unknown protein (e.g., encoded by a gene cloned from a cDNA library), or a random polypeptide sequence (or be a random sequence included in a larger polypeptide sequence).

25 To isolate DNA sequences encoding novel interacting proteins, members of a DNA expression library (e.g., a cDNA or synthetic DNA library, either random or intentionally biased) can be fused in-frame to the activation tag (e.g., the polymerase interaction domain or activation domain) to generate a variegated library of prey fusion proteins. Those library-encoded proteins that physically interact with the promoter-bound bait fusion protein

- 25 -

detectably activate expression of the reporter gene and provide a ready assay for identifying a particular DNA clone encoding an interacting protein of interest.

In an exemplary embodiment, cDNAs may be constructed from any mRNA population and inserted into an equivalent expression vector. Such a library of choice may 5 be constructed de novo using commercially available kits (e.g., from Stratagene, La Jolla, CA) or using well established preparative procedures (see, for example, Current Protocols in Molecular Biology, Eds. Ausubel et al. John Wiley & Sons: 1992). Alternatively, a number of cDNA libraries (from a number of different organisms) are publicly and commercially available; sources of libraries include, e.g., Clontech (Palo Alto, CA) and 10 Stratagene (La Jolla, CA). It is also noted that prey polypeptide need not be naturally occurring full-length proteins. In preferred embodiments, prey proteins are encoded by synthetic DNA sequences, are the products of randomly generated open reading frames, are open reading frames synthesized with an intentional sequence bias, or are portions thereof. Preferably, such short randomly generated sequences encode peptides between, for 15 example, 4 and 60 amino acids in length.

It will be appreciated by those skilled in the art that many variations of the prey and bait fusion proteins can be constructed and should be considered within the scope of the present invention. For example, it will be understood that, for screening polypeptide libraries, the identity of the prey polypeptide can be fixed and the bait protein can be varied 20 to generate the library. Indeed, in certain embodiments it will be desirable to derive the prey fusion protein with a fixed prey polypeptide rather than a variegated library on the grounds that the single prey fusion protein can be easily tested for its ability to be assembled into a functional RNA polymerase enzyme. Moreover, where the prey fusion protein is derived with a polymerase interaction domain, the bait fusion protein is likely to be less 25 sensitive to variations caused by the different peptides of the library than is the prey fusion protein. In such embodiments, a variegated bait polypeptide library can be used to create a library of bait fusion proteins to be tested for interaction with a particular prey protein.

While it will generally be desirable for the DBD and bait polypeptide portions of the bait fusion protein, and activation tag and prey polypeptide portions of the prey fusion 30 protein to be derived from different, e.g., heterologous, proteins, the present invention also contemplates embodiments of the instant assay wherein one of the two bait or prey proteins

- 26 -

is a naturally occurring protein rather than a heterologous fusion protein. As an illustration, the bait protein can be a dimeric transcriptional activator which undergoes a higher order tetramerization reaction. That dimer-dimer interaction can be selected as the target of an assay to identify an agent which selectively disrupts the inter-dimer contacts. In such 5 embodiments, the full-length transcriptional activator can serve the role of the bait protein, and the prey fusion protein can include, for example, that portion of the transcriptional activator which is involved in the formation of tetrameric complexes.

Moreover, either or both the prey and bait proteins, if desired, may include epitope tags (e.g., portions of the c-myc protein or the flag epitope available from Immunex). The 10 epitope tag can facilitate a simple immunoassay for fusion protein expression, e.g. to detect the presence and folding of the fusion protein.

In other embodiments of the subject ITS, particularly those in which a polypeptide library is displayed on either the bait or prey protein, the fusion proteins can be generated to include, in addition to the test polypeptide sequences, a polypeptide sequence with another 15 known polypeptide sequence. Thus, a prey fusion protein can be generated having the following exemplary formula: A-B-C, where A is an  $\alpha$ -NTD, B is a control binding sequence (such as the C terminal domain [CTD] of  $\lambda$ cI), and C is the test polypeptide sequence. To assure oneself that the fusion protein is correctly folded, the fusion protein can be first tested in an ITS using  $\lambda$ cI CTD in the bait protein --the C terminal domain 20 included in the prey protein providing a means for binding (by dimerization) with the bait. Prey fusion proteins which pass this control ITS can then be sampled in an ITS wherein bait is constructed with test polypeptide(s). Of course it will be appreciated that the order of the control and test polypeptides can be reversed.

In other embodiments, the construct encoding the prey (or bait) fusion protein can 25 include a promoter for *in vitro* translation (e.g., a T7 promoter) of the target polypeptide. c.f., Yavuzer et al. (1995) *Gene* 165:93. Such constructs can be used to eliminate subcloning steps necessary to carry out certain validation assays often undertaken after the initial identification of the protein in the interaction trap, e.g., to determine if the binding of 30 the two hybrid proteins is truly the result of an interaction between the bait and prey polypeptides *per se*.

- 27 -

In another aspect of the present invention, the DNA sequence encoding the prey protein (or alternatively the bait protein) is embedded in a DNA sequence encoding a conformation-constraining protein (i.e., a protein that decreases the flexibility of the amino and carboxy termini of the prey protein). Such embodiments are preferred where the prey 5 polypeptide is a relatively short peptide, e.g., 5-25 amino acid residues. In general, conformation-constraining proteins act as scaffolds or platforms, which limit the number of possible three dimensional configurations the peptide or protein of interest is free to adopt. Preferred examples of conformation-constraining proteins are thioredoxin or other thioredoxin-like sequences, but many other proteins are also useful for this purpose. 10 Preferably, conformation-constraining proteins are small in size (generally, less than or equal to 200 amino acids), rigid in structure, of known three dimensional configuration, and are able to accommodate insertions of proteins of interest without undue disruption of their structures. A key feature of such proteins is the availability, on their solvent exposed surfaces, of locations where peptide insertions can be made (e.g., the thioredoxin active-site 15 loop).

As mentioned above, one preferred conformation-constraining protein according to the invention is thioredoxin or other thioredoxin-like proteins. The three dimensional structure of *E. coli* thioredoxin is known and contains several surface loops, including a distinctive Cys-Cys active-site loop between residues Cys33 and Cys36 which protrudes 20 from the body of the protein. This Cys-Cys active-site loop is an identifiable, accessible surface loop region and is not involved in interactions with the rest of the protein which contribute to overall structural stability. It is therefore a good candidate as a site for prey protein insertions. Both the amino- and carboxyl-termini of *E. coli* thioredoxin are on the surface of the protein and are also readily accessible for fusion construction.

25 It may be preferred for a variety of reasons that prey (or bait) polypeptides be fused within the active-site loop of thioredoxin or thioredoxin-like molecules. The face of thioredoxin surrounding the active-site loop has evolved, in keeping with the protein's major function as a nonspecific protein disulfide oxido-reductase, to be able to interact with a wide variety of protein surfaces. The active-site loop region is found between segments of strong 30 secondary structure and this provides a rigid platform to which one may tether prey proteins. A small prey protein inserted into the active-site loop of a thioredoxin-like protein

is present in a region of the protein which is not involved in maintaining tertiary structure. Therefore the structure of such a fusion protein is stable. Thus, relatively short peptides may be displayed as part of the prey fusion protein by virtue of the fusion of the thioredoxin protein to a polymerase interaction domain. Such embodiments are useful for screening 5 peptide libraries for interactors with a particular target bait protein.

The subject assay can also be used to generate antibody equivalents for specific determinants, e.g., such as single chain antibodies, minibodies or the like. Indeed, the subject method can be used to identify a novel binding partner for a given epitope/determinant where the new binding partner is a completely artificial polypeptide. 10 For example, a target polypeptide (or epitope thereof) for which an antibody or antibody equivalent is sought can be displayed on either the bait or prey fusion protein. A library of potential binding partners can be arrayed on the other fusion protein, as appropriate. Interactions between the target polypeptide and members of the library of binding partners can be detected according to methods described herein. Thus, the present invention 15 provides a convenient method for identifying recombinant nucleic acid sequences which encode proteins useful in the replacement of, e.g., monoclonal antibodies.

In another embodiment of the subject ITS, the system can be used to identify proteolytic activities which cleave a given polypeptide sequence, or to identify the sequence specificity for a given protease. For example, in the embodiment of the subject ITS 20 illustrated in Figure 1B, a desired cleavage sequence can be introduced into the bait or prey fusion proteins such that, upon cleavage of the fusion protein at that sequence, the DNA localization of the prey protein is lost. To further illustrate, a substrate sequence for a proteolytic activity is desired can be engineered into the linker sequence separating the N- and C-terminal domains of the bait protein shown in Figure 1B. In the absence of 25 proteolysis of that sequence, the intact prey and bait proteins induce expression of a reporter gene (or "inverter" gene as appropriate). The presence in the cell of a proteolytic activity which recognizes the substrate sequence can result in cleavage of the bait protein, separating the DBD from that portion of the protein which interacts with the prey fusion protein. Such embodiments of the ITS can be used to screen libraries of proteolytic proteins, e.g., derived 30 from cDNA libraries, catalytic antibodies, or generated by combinatorial mutagenesis of existing enzymes.

- 29 -

In other embodiments, peptide libraries can be engineered into one of the fusion proteins and proteolysis of the fusion protein by a predetermined proteolytic activity used to identify the sequence specificity of the proteolytic activity and/or optimize the sequence for a substrate or inhibitor for the proteolytic activity. For example, a variety of proteases have 5 been identified as being involved in various disease states. In many instances, the substrate specificity for a protease has not yet been fully determined or optimized. Utilizing the subject ITS, the substrate specificity for a given protease can be accurately determined, and selective substrates or inhibitors, as appropriate, can be developed based on that sequence information.

10 In still other embodiments, the subject ITS can be derived to score for heteromeric combinations of three or more proteins by providing two or more different bait fusion proteins and/or two or more different prey fusion proteins in the same system, i.e., at least three different fusion proteins. This concept is illustrated by an example using  $\alpha$ -NTD fusion proteins.

15 The  $\alpha$  subunit of *E. coli* RNA polymerase plays a key role in assembly of the core enzyme. In previous studies, it has been demonstrated that the holoenzyme includes two  $\alpha$  subunits, only one of which interacts with  $\beta$ . Assembly-deficient mutants of  $\alpha$  have been identified, such as  $\alpha$ -R45A (having substituted Ala for Arg at residue 45). This mutant dimerizes, but does not assemble  $\beta$  subunits. See Kimura et al. (1995) *J Mol Biol* 254:342.

20 When over-expressed in cells also expressing wildtype  $\alpha$ , the equilibrium of the system favors formation of holoenzyme complexes which are heterologous with respect to  $\alpha$ , e.g., including one wildtype and one R45A mutant subunit. Thus, making fusion proteins with a DNA binding domain, and with each of the wildtype and R45A N-NTDs, the system can accommodate three different polypeptide sequences which can be tested for simultaneous 25 interactions. In other embodiments, fusing the same polypeptide sequence to the two different  $\alpha$ -NTD sequences can be used to distinguish oligomerization mechanisms, e.g., distinguish tetramerization from pairwise dimerization.

*V. Reporter gene constructs*

- 30 -

The reporter gene of this invention ultimately measures the end stage of the above described cascade of events, e.g., transcriptional modulation, and, if desired, permits the isolation of ITS cells on the basis of that criteria. Accordingly, in practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order

5 to generate a detection signal dependent on interaction of the bait and prey fusion proteins. Typically, the reporter gene construct will include a reporter gene in operative linkage with one or more transcriptional regulatory elements which include, or are linked to, a DBD recognition element for the DBD of the bait fusion protein, with the level of expression of the reporter gene providing the prey protein interaction-dependent detection signal. Many

10 reporter genes and transcriptional regulatory elements useful in the subject flow-ITS are known to those of skill in the art and others may be readily identified or synthesized. Moreover, DBD recognition elements are known in the art for a wide variety of DNA binding domains which may be used to construct the bait proteins of the present invention. Exemplary recognition elements include the  $\lambda$  operator, the LexA operator, the pho box,

15 and the like.

A "reporter gene" includes any gene that expresses a detectable gene product, which may be RNA or protein. Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable

20 properties.

Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol. Cell. Biol. 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984),

25 PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); phycobiliproteins (especially phycoerythrin); green fluorescent protein (GFP: see Valdivia et al. (1996) Mol Microbiol 22: 367-78; Cormack et al. (1996) Gene 173 (1 Spec No): 33-8; and Fey et al. (1995) Gene 165:127-130; alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), secreted alkaline

30 phosphatase (Cullen and Malim (1992) Methods in Enzymol. 216:362-368). Other examples of suitable reporter genes include those which encode proteins conferring

- 31 -

drug/antibiotic resistance to the host bacterial cell, or which encode proteins required to complement an auxotrophic phenotype. A preferred reporter gene is the *spc* gene, which confers resistance to spectinomycin.

The amount of transcription from the reporter gene may be measured using any 5 method known to those of skill in the art to be suitable. For example, specific mRNA expression may be detected using Northern blots or specific protein product may be identified by a characteristic stain or an intrinsic activity.

In preferred embodiments, the gene product of the reporter is detected by an intrinsic 10 activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detection signal based on color, fluorescence, or luminescence.

The amount of expression from the reporter gene is then compared to the amount of expression in either the same cell in the absence of the test compound or it may be compared with the amount of transcription in a substantially identical cell that lacks 15 heterologous DNA, such as the gene encoding the prey fusion protein. Any statistically or otherwise significant difference in the amount of transcription indicates that the prey fusion protein interacts with the bait fusion protein.

In other preferred embodiments, the reporter or marker gene provides a selection 20 method such that cells in which the reporter gene is activated have a growth advantage. For example the reporter could enhance cell viability, e.g., by relieving a cell nutritional requirement, and/or provide resistance to a drug. For example the reporter gene could encode a gene product which confers the ability to grow in the presence of a selective agent, e.g., chorlamphenicol or kanamycin.

In bacteria, suitable positively selectable (beneficial) genes include genes involved 25 in biosynthesis or drug resistance. Countless other genes are potential selective markers. Certain of the above are involved in well-characterized biosynthetic pathways. In the simplest case, the cell is auxotrophic for an amino acid, such as histidine (requires histidine for growth), in the absence of activation of the reporter gene. Activation leads to synthesis of an enzyme required for biosynthesis of the amino acid and the cell becomes prototrophic

- 32 -

for that amino acid (does not require an exogenous source). Thus the selection is for growth in the absence of that amino acid in the culture media.

Another class of useful reporter genes encode cell surface proteins for which antibodies or ligands are available. Expression of the reporter gene allows cells to be 5 detected or affinity purified by the presence of the surface protein.

In appropriate assays, so-called counterselectable or negatively selectable genes may be used.

The marker gene may also be a screenable gene. The screened characteristic may be a change in cell morphology, metabolism or other screenable features. Suitable markers 10 include  $\beta$ -galactosidase, alkaline phosphatase, horseradish peroxidase, luciferase, bacterial green fluorescent protein,; secreted alkaline phosphatase (SEAP); and chloramphenicol transferase (CAT). Some of the above can be engineered so that they are secreted (although not  $\beta$ -galactosidase). A preferred screenable marker gene is  $\beta$ -galactosidase; bacterial cells expressing the enzyme convert the colorless substrate Xgal into a blue pigment.

15 In general, many of the embodiments of the ITS described above rely upon expression the reporter as a positive readout, typically manifested either (1) as an enzyme activity (e.g.,  $\beta$ -galactosidase) or (2) as enhanced cell growth on a defined medium (e.g., antibiotic resistance). Thus, these methods are suited for identifying a positive interaction of the bait and prey polypeptides, but are not well suited for identifying agents or conditions 20 which inhibit intermolecular association between two polypeptide sequences. In part, this is because a failure to obtain expression of the reporter gene can result from many events which do not stem from a specific inhibition of binding of the two hybrid proteins. For example, an ITS using a reporter gene that stimulates growth under defined conditions theoretically can be used to screen for agents that inhibit the intermolecular association of 25 the two hybrid proteins, but it will be difficult or impossible to discriminate agents that specifically inhibit the association of the two hybrid proteins from agents which simply inhibit cell growth. Thus, an agent which is cytotoxic to the bacterial cell will prevent cell growth without specifically inhibiting the interaction of two hybrid proteins and will score falsely as a positive hit. Similarly, an ITS using a lacZ reporter gene or the like, or a 30 cytotoxic gene, will falsely score general transcription or translation inhibitors as being

- 33 -

inhibitors of two hybrid protein binding. Thus, ITS embodiments that produce a positive readout contingent upon intermolecular binding of the bait and prey proteins are generally not suitable for screening for agents which inhibit binding of the two hybrid proteins.

To avoid such confounding results, the ITS format can be modified slightly to 5 provide a "reverse ITS". In the reverse ITS, the reporter gene encodes a transcriptional repressor which is expressed upon interaction of the bait and prey proteins. However, the host cell also includes a second reporter gene which, but for an operator sequence responsive to the repressor protein produced by the first reporter gene, would otherwise be expressed. Thus, the gene product of the first reporter gene regulates expression of the 10 second reporter gene, the expression of the latter provides a means for indirectly scoring for the expression of the former. Essentially, the first reporter gene can be seen as a signal inverter.

In this exemplary system, the bait and prey proteins positively regulate expression of 15 the first reporter gene. Accordingly, where the first reporter gene is a repressor of expression of the second reporter gene, relieving expression of the first reporter gene by inhibiting the formation of complexes between the bait and prey proteins concomitantly relieves inhibition of the second reporter gene. For example, the first reporter gene can include the coding sequences for  $\lambda$ cI. The second reporter gene can accordingly be a 20 positive signal, such as providing for growth (e.g., drug selection or auxotrophic relief), and is under the control of a promoter which is constitutively active, but can be repressed by  $\lambda$ cI. In the absence of an agent which inhibits the interaction of the bait and prey protein, the  $\lambda$ cI protein is expressed. In turn, that protein represses expression of the second reporter gene. However, an agent which disrupts binding of the bait and prey proteins results in a 25 decrease in  $\lambda$ cI expression, and consequently an increase in expression of the second reporter gene as  $\lambda$ cI repression is relieved. Hence, the signal is inverted.

In yet another embodiment for detecting agents which disrupt the bait-prey interaction, it is envisioned that under certain conditions the interaction between bait and prey fusion proteins might result in transcription repression rather than activation. For example, it is speculated that sufficiently strong binding between a bait fusion protein and a 30 prey fusion protein may impede the escape of the polymerase from the promoter, which escape is required for elongation of a transcript, thus repressing transcription. In particular,

- 34 -

a strong interaction between the bait and protein proteins, combined with a strong promoter (e.g., one which is more efficient at binding the polymerase complex even in the absence of transcription factors) can result in repression of reporter gene expression. Under these conditions an inhibitor of bait-prey complex formation will, over a certain concentration range, cause the effective association constant of the complex to be reduced sufficiently to result in relief of the repression and concomitant transcription of the reporter gene. At higher concentrations, inhibitors of the bait-prey complex may result in inhibition (or return to basal levels) of transcription by the loss of bait-prey complexes. Thus, in one embodiment, the candidate agent can be spotted on a lawn of reagent cells plated on a solid media. The diffusion of the candidate agent through the solid medium surrounding the site at which it was spotted will create a diffusional effect. For agents which inhibit the formation of bait-prey complexes, a halo of reporter gene expression would be expected in an area which corresponds to concentrations of the agent which offset the effect of the repression due to strong association between the two hybrid proteins, but which are not so great as to substantially inhibit the formation of bait-prey complexes.

Still another consideration in generating the reporter gene construct concerns the placement of the DBD recognition element relative to the reporter gene and other transcriptional elements with which it is associated. In most embodiments, it will be desirable to position the recognition element at an inert position. In some instances, the axial position of the DBD relative to the promoter sequences can be important.

In certain embodiments, the sensitivity of the ITS can be enhanced for detecting weak protein-protein interactions by placing the DBD recognition sequence at a position permitting secondary interactions (if any) between other portions of the bait fusion protein and the RNA polymerase complex. For example, as described in the appended examples, an apparent synergistic effect was observed when the  $\lambda$  operator was moved close to or at its normal position. While not wishing to be bound by any particular theory, this synergism is speculated to be the result of a bait-prey interaction and second interaction between DBD of  $\lambda$ cI and a second polymerase subunit ( $\sigma$ ).

It will also be understood by those skilled in the art that the sensitivity to the strength of the interactions between the bait and prey proteins can be "tuned" by adjusting the sequence of the recognition element. For example, the use of a strong  $\lambda$  operator instead

- 35 -

of weak can improve the sensitivity of the assay to weak bait-prey interactions, as well as help to overcome lack of dimerization if no dimerization signals are included in the bait fusion protein.

In particular embodiments, it may desirable to provide two or more reporter gene constructs which are regulated by interaction of the bait and prey proteins. The simultaneous expression of the various reporter genes (whether provided on the same or separate plasmids) provides a means for distinguishing actual interaction of the bait and prey proteins from, e.g., mutations or other spurious activation of the reporter gene.

#### VI. Host cells

10 Exemplary prokaryotic host cells are gram-negative bacteria such as *Escherichia coli*, or gram-positive bacteria such as *Bacillus subtilis*.

Recognized prokaryotic hosts include bacterial strains of *Escherichia*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, *Shigella* and the like. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

15 Preferred prokaryotic host cells for use in carrying out the present invention are strains of the bacteria *Escherichia*, although *Bacillus* and other genera are also useful. Techniques for transforming these hosts and expressing foreign genes cloned in them are well known in the art (see e.g., Maniatis et al. and Sambrook et al., *ibid.*). Vectors used for expressing foreign genes in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter which functions in the host cell. Appropriate promoters including *trp* (Nicholset al. (1983) *Meth. Enzymol.* 101:155-164), *lac* (Casadaban et al. (1980) *J. Bacteriol.* 143:971-980), and phage gamma promoter systems (Queen (1983) *J. Mol. Appl. Genet.* 2:1-10). Plasmids useful for transforming bacteria include pBR322 (Bolivar et al. (1977) *Gene* 2:95-113), the pUC plasmids (Messing 20 (1983) *Meth. Enzymol.* 101:20-77), Vieira and Messing (1982) *Gene* 19:259-268), pCQV2 (Queen, *supra*), pACYC plasmids (Chang et al. (1978) *J. Bacteriol.* 134:1141), pRW 25 plasmids (Lodge et al. (1992) *FEMS Microbiol Lett* 95:271), and derivatives thereof.

30 The choice of appropriate host cell will also be influenced by the choice of detection signal. For instance, reporter constructs, as described below, can provide a selectable or screenable trait upon transcriptional activation (or inactivation). The reporter gene may be

- 36 -

an unmodified gene already in the host cell pathway, such as sporulation genes. It may be a host cell gene that has been operably linked to a "bait-responsive" promoter. Alternatively, it may be a heterologous gene that has been so linked. Suitable genes and promoters are discussed above. Accordingly, it will be understood that to achieve selection or screening, 5 the host cell must have an appropriate phenotype. For example, introducing a histidine biosynthesis gene into a yeast that has a wild-type form of that gene would frustrate genetic selection. Thus, to achieve nutritional selection, an auxotrophic strain will be desired which is complemented by expression of the reporter gene.

In other embodiments, the host cell can be a eukaryotic cell, particularly a yeast cell, 10 which has been engineered to express a sufficient number of the bacterial polymerase subunits necessary to induce (reporter) gene expression in the cell in a manner dependent on the bait and prey proteins and the bacterial RNA polymerase subunits. It may be desirable in such embodiments to include a nuclear localization signal as part of one or more of the bacterial proteins. Regulatory sequences for the recombinant expression of these proteins in 15 eukaryotic cells may also need to be optimized.

#### *VII. Exemplary Uses of the Prokaryotic ITS*

The prokaryotic ITS of the present invention can be used, *inter alia*, for identifying protein-protein interactions, e.g., for generating protein linkage maps, for 20 identifying therapeutic targets, and/or for general cloning strategies. As described above, the ITS can be derived with a cDNA library to produce a variegated array of bait or prey proteins which can be screened for interaction with, for example, a known protein expressed as the corresponding fusion protein in the ITS. In other embodiments, both the bait and prey proteins can be derived to each provide variegated libraries of polypeptide sequences. 25 One or both libraries can be generated by random or semi-random mutagenesis. For example, random libraries of polypeptide sequences can be "crossed" with one another by simultaneous expression in the subject assay. Such embodiments can be used to identify novel binding pairs of polypeptides.

Alternatively, the subject ITS can be used to map residues of a protein involved in a 30 known protein-protein interaction. Thus, for example, various forms of mutagenesis can be

- 37 -

utilized to generate a combinatorial library of either bait or prey polypeptides, and the ability of the corresponding fusion protein to function in the ITS can be assayed. Mutations which result in diminished (or potentiated) binding between the bait and prey fusion proteins can be detected by the level of reporter gene activity. For example, mutants of a 5 particular protein which alter interaction of that protein with another protein can be generated and isolated from a library created, for example, by alanine scanning mutagenesis and the like (Ruf et al., (1994) Biochemistry 33:1565-1572; Wang et al., (1994) J. Biol. Chem. 269:3095-3099; Balint et al., (1993) Gene 137:109-118; Grodberg et al., (1993) Eur. J. Biochem. 218:597-601; Nagashima et al., (1993) J. Biol. Chem. 268:2888-2892; Lowman 10 et al., (1991) Biochemistry 30:10832-10838; and Cunningham et al., (1989) Science 244:1081-1085), by linker scanning mutagenesis (Gustin et al., (1993) Virology 193:653-660; Brown et al., (1992) Mol. Cell Biol. 12:2644-2652; McKnight et al., (1982) Science 232:316); by saturation mutagenesis (Meyers et al., (1986) Science 232:613); by PCR mutagenesis (Leung et al., (1989) Method Cell Mol Biol 1:11-19); or by random 15 mutagenesis (Miller et al., (1992) A Short Course in Bacterial Genetics, CSHL Press, Cold Spring Harbor, NY; and Greener et al., (1994) Strategies in Mol Biol 7:32-34). Linker scanning mutagenesis, particularly in a combinatorial setting, is an attractive method for identifying truncated (bioactive) forms of a protein, e.g., to establish binding domains.

In other embodiments, the ITS can be designed for the isolation of genes encoding 20 proteins which physically interact with a protein/drug complex. The method relies on detecting the reconstitution of a transcriptional activator in the presence of the drug, such as rapamycin, FK506 or cyclosporin. If the bait and prey fusion proteins are able to interact in a drug-dependent manner, the interaction may be detected by reporter gene expression.

Another aspect of the present invention relates to the use of the prokaryotic ITS in 25 the development of assays which can be used to screen for drugs which are either agonists or antagonists of a protein-protein interaction of therapeutic consequence. In a general sense, the assay evaluates the ability of a compound to modulate binding between the bait and prey polypeptides. Exemplary compounds which can be screened include peptides, nucleic acids, carbohydrates, small organic molecules, and natural product extract libraries, 30 such as isolated from animals, plants, fungus and/or microbes.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. The subject ITS-derived screening assays can be carried out in such a format, and accordingly may be used as a "primary" screen.

5 Accordingly, in an exemplary screening assay of the present invention, an ITS is generated to include specific bait and prey fusion proteins known to interact, and compound(s) of interest. Detection and quantification of reporter gene expression provides a means for determining a compound's efficacy at inhibiting (or potentiating) interaction between the bait and prey polypeptides. In certain embodiments, the approximate efficacy of the

10 compound can be assessed by generating dose response curves from reporter gene expression data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, expression of the reporter gene is quantitated in the absence of the test compound.

In an illustrative embodiment, the ITS assay can be used to identify cyclophilin or

15 rapamycin mimetics by screening for agents which potentiate the interaction of an FK506 binding protein (FKBP) and a cyclophilin or TOR1 protein. For example, rapamycin-like drugs can be identified by the present invention which have enhanced tissue-type or cell-type specificity relative to rapamycin. The identification of such compounds can be enhanced by the use of differential screening techniques which detect and compare drug-

20 mediated formation of two or more different types of FKBP/cyclophilin or FKBP/TOR complexes. To further illustrate, by side-by-side comparison of assays generated with mammalian and yeast proteins, the subject ITS can be used to identify rapamycin mimetics which preferentially inhibit proliferation of yeast cells or other lower eukaryotes, but which have a substantially reduced effect on mammalian cells, thereby improving therapeutic

25 index of the drug as an anti-mycotic agent relative to rapamycin.

In another exemplary embodiment, a therapeutic target devised as the bait-prey complex is contacted with a peptide library with the goal of identifying peptides which potentiate or inhibit the bait-prey interaction. Many techniques are known in the art for expression peptide libraries intracellularly. In one embodiment, the peptide library is

30 provided as part of a chimeric thioredoxin protein, e.g., expressed as part of the active loop (supra).

- 39 -

In yet another embodiment, the bacterial ITS can be generated in the form of a diagnostic assay to detect the interaction of two proteins, e.g., e.g., where the gene from one is isolated from a biopsied cell. For instance, there are many instances where it is desirable to detect mutants which, while expressed at appreciable levels in the cell, are defective at 5 binding other cellular proteins. Such mutants may arise, for example, from fine mutations, e.g., point mutants, which may be impractical to detect by the diagnostic DNA sequencing techniques or by the immunoassays. The present invention accordingly further contemplates diagnostic screening assays which generally comprise cloning one or more cDNAs from a sample of cells, and expressing the cloned gene(s) as part of an ITS under 10 conditions which permit detection of an interaction between that recombinant gene product and a target protein. Accordingly, the present invention provides a convenient method for diagnostically detecting mutations to genes encoding proteins which are unable to physically interact with a target "bait" protein, which method relies on detecting the reconstitution of a transcriptional activator in a bait/prey-dependent fashion.

15 To illustrate, the subject ITS can be used to detect inactivating mutations of the CDK4/p16<sup>INK4a</sup> interaction. Recent discoveries have brought several cell-cycle regulators into sharp focus as factors in human cancer. Among the most conspicuous types of molecule to emerge from ongoing studies in this field are the cyclin-dependent kinase inhibitors such as p16. (Serrano et al. (1993) *Nature* 366:704; and Okamoto et al. (1994) 20 *PNAS* 91:11045) The p16 protein has several hallmarks of a tumor suppressor and is perfectly positioned to regulate critical decisions in cell growth. The p16 gene appears to be a particularly significant target for mutation in sporadic tumors and in at least one form of hereditary cancer. In an exemplary embodiment of the diagnostic ITS, a first hybrid gene 25 comprises the coding sequence for a DNA-binding domain fused in frame to the coding sequence for a bait protein, e.g., CDK4 or CDK6. The second hybrid protein encodes a polymerase interaction domain fused in frame to a gene encoding the sample protein, e.g. a p16 gene (cDNA) amplified from a cell sample of a patient. If the bait and sample proteins are able to interact, e.g., form a CDK/p16 complex, then RNA polymerase is recruited to the promoter of a reporter gene which is operably linked to a DBD recognition element, thereby 30 causing expression of the reporter gene.

- 40 -

Moreover, it will be apparent that the subject two hybrid assay can be used generally to detect mutations in other cellular proteins which disrupt protein-protein interactions. For example, it has been shown that the transcription factor E2F-4 is bound to the p130 pocket protein, and that such binding effectively suppresses E2F-4-mediated trans-activation required for control of G<sub>0</sub>/G<sub>1</sub> transition. Mutants which result in disruption of this interaction can be detected in the subject assay.

Similarly, Rb and Rb-like proteins (such as p107) act to control cell-cycle progression through the formation of complexes with several cellular proteins. In fact, a recent article concerning familial retinoblastoma has reported a new class of Rb mutants found in retinal lesions, which mutants were defective in protein binding ("pocket") activity (see, for example, Kratzke et al. (1994) *Oncogene* 9:1321-1326). Moreover, mutant forms of c-myc have been demonstrated in various lymphomas, e.g., Burkitt lymphomas, which mutants are resistant to p107-mediated suppression. Accordingly, the diagnostic two hybrid assay of the present invention can be used to detect mutations in Rb or Rb-like proteins which disrupt binding to other cellular proteins, e.g., myc, E2F, c-Abl, or upstream binding factor (UBF), or vice-versa.

In another embodiment, the subject diagnostic assay can be employed to detect mutations which disrupt binding of the p53 protein with other cellular proteins, as for example, the Wilm's tumor suppresser protein WT1. Recent observations by Maheswaran et al. (1993, *PNAS* 90:5100-5104) have demonstrated that p53 can physically interact with WT1, and that this interaction modulates the ability of each protein to transactivate their respective targets. In fact, in contrast to the proposed function of *WT1* as a transcriptional repressor, potent transcriptional activation by *WT1* of reporter genes driven by EGR1 in cells lacking wild type p53 indicates that transcriptional repression is not an intrinsic property of *WT1*. Instead, transcriptional repression by *WT1* may result from its interaction with p53. Accordingly, mutations in p53 which do not effect the cellular concentration of this protein, but which rather down regulate its ability to bind to and repress *WT1*, may give rise to Wilm's tumors, and other disease states associated with deregulation of *WT1*.

In still another embodiment, the diagnostic two hybrid assay can be used to detect mutations in pairs of signal transduction proteins. For example, the present assay can be

used to detect mutations in the *ras* protein or other cellular proteins which interact with *ras*, e.g., *ras* GTPase activating proteins (GAPs).

The method of the present invention, as described above, may be practiced using a kit for detecting interaction between a target protein and a sample protein. In an illustrative 5 embodiment, the kit includes two vectors, a host cell, and (optionally) a set of primers for cloning one or more target proteins from a patient sample. The first vector may contain a promoter, a transcription termination signal, and other transcription and translation signals functionally associated with the first chimeric gene in order to direct the expression of the first chimeric gene. The first chimeric gene includes a DNA sequence that encodes a DNA-10 binding domain and a unique restriction site(s) for inserting a DNA sequence encoding the target protein or protein fragment in such a manner that the target protein is expressed as part of a hybrid protein with the DNA-binding domain. The first vector also includes a means for replicating itself (e.g., an origin of replication) in the host cell. In preferred embodiments, the first vector also includes a first marker gene, the expression of which in 15 the host cell permits selection of cells containing the first marker gene from cells that do not contain the first marker gene. Preferably, the first vector is a plasmid.

The kit also includes a second vector which contains a second chimeric gene. The second chimeric gene also includes a promoter and other relevant transcription and translation sequences to direct expression of the prey fusion protein. The second chimeric 20 gene also includes a DNA sequence that encodes a polymerase interaction domain (or an activation domains) and a unique restriction site(s) to insert a DNA sequence encoding the sample protein, or fragment thereof, into the vector in such a manner that the target protein is capable of being expressed as part of a hybrid protein with the polymerase interaction domain.

25 In general, the kit will also be provided with one of the two vectors already including the bait protein. For example, the kit can be configured for detecting mutations to a p16-gene which result in loss of binding to CDK4. Accordingly, the first vector could be provided with a CDK4 open reading frame fused in frame to the DNA-binding domain to provide a CDK4 bait protein. p16-gene open reading frames can be cloned from a cell 30 sample and ligated into the second vector in frame with the polymerase interaction domain.

- 42 -

Where the kit also provides primers for cloning a p16-gene into the two hybrid assay vectors, the primers will preferably include restriction endonuclease sites for facilitating ligation of the amplified gene into the insertion site flanking the DNA-binding domain or activating domain.

5 Accordingly in using the kit, the interaction of the target protein and the sample protein in the host cell causes a measurably greater expression of the reporter gene than when the DNA-binding domain and the polymerase interaction domain are present in the absence of an interaction between the two fusion proteins. The cells containing the two hybrid proteins are incubated in/on an appropriate medium and the cells are monitored for  
10 10 the measurable activity of the gene product of the reporter construct. A positive test for this activity is an indication that the target protein and the sample protein have interacted. Such interaction brings their respective DNA-binding and polymerase interaction domain into sufficiently close proximity to cause efficient transcription of the reporter gene.

15

#### *Exemplification*

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention.

20 The C-terminal domain of the alpha subunit of RNA polymerase ( $\alpha$ -CTD) mediates the effects of many transcriptional activators in bacteria, likely through direct contact. The  $\alpha$ -CTD was replaced with the C-terminal domain of the bacteriophage  $\lambda$  repressor, a domain that forms dimers and higher order oligomers. It is then demonstrated that an artificial promoter bearing a single  $\lambda$  operator in its upstream region is activated by  $\lambda$   
25 repressor in cells that express the hybrid  $\alpha$  gene. The following examples further show that mutations in  $\lambda$  repressor that weaken the CTD oligomerization interaction also decrease activation in the strain bearing the hybrid  $\alpha$  gene. These findings show that the strength of an arbitrary protein-protein interaction determines the magnitude of gene activation. Thus, for at least certain promoters, recruitment of RNA polymerase to the DNA is sufficient for  
30 gene activation.

- 43 -

RNA polymerase in *E. coli* consists of an enzymatic core composed of subunits  $\alpha$ ,  $\beta$ , and  $\beta'$  in the stoichiometry  $\alpha_2\beta\beta'$ , and one of several alternative  $\sigma$  factors responsible for specific promoter recognition. The  $\alpha$  subunit, which initiates the assembly of RNA polymerase by forming a dimer, has two independently folded domains. The larger amino-  
5 terminal domain ( $\alpha$ -NTD) mediates dimerization and the subsequent assembly of polymerase. The carboxy-terminal domain ( $\alpha$ -CTD), which is tethered to the  $\alpha$ -NTD by a flexible linker region, interacts with a DNA sequence known as the "UP-element" that is found upstream of the -35 region of certain particularly strong promoters. The  $\alpha$ -CTD is also the target of action of a large class of transcriptional activators.

10 The Cyclic AMP Receptor Protein (CRP) is the most intensively studied example of a transcriptional activator that exerts its effect on the  $\alpha$ -CTD. Several lines of evidence indicate that CRP uses a well-defined activating region consisting of a nine amino acid surface-exposed loop to contact the  $\alpha$ -CTD directly when bound to its recognition site (centered at position -61.5) upstream of the familiar *lac* promoter. In the case of CRP as  
15 well as several other activators, specific amino acid residues in the  $\alpha$ -CTD have been identified that are required for activation. The available evidence suggests that activation by this class of activators involves direct contact with one or another target region on the  $\alpha$ -CTD. However, this evidence does not establish whether the  $\alpha$ -CTD plays some special role or whether any protein-protein contact would suffice.

20 To address this question, the natural interaction between activator and  $\alpha$ -CTD was replaced with a different interaction involving a protein domain that does not ordinarily mediate transcriptional activation. To do this, the well-defined properties of the C-terminal domain (CTD) of the bacteriophage  $\lambda$  repressor were relied upon.

25 The  $\lambda$  repressor ( $\lambda$ cI) is a two-domain protein that functions as both a repressor and an activator of transcription.  $\lambda$ cI binds DNA as a dimer, and pairs of dimers bind cooperatively to adjacent operator sites (Figure 1A). The N-terminal domain contacts the DNA and interacts with RNA polymerase when  $\lambda$ cI is bound at promoter  $P_{RM}$ , whereas the CTD mediates both dimer formation and the dimer-dimer interaction that results in cooperativity. A large number of  $\lambda$ cI mutants specifically defective for cooperative  
30 binding to DNA have been isolated and these mutants bear single amino acid substitutions in the CTD.

- 44 -

It was reasoned that if the  $\alpha$ -CTD was replaced with the  $\lambda$ cI-CTD, the resulting  $\alpha$ -cI fusion protein would display a dimeric target that could be contacted by an appropriately positioned  $\lambda$ cI dimer (Figure 1B). This would test whether the same protein-protein interaction that ordinarily mediates the cooperative binding of pairs of  $\lambda$ cI dimers to the  
5 DNA would mediate transcriptional activation when the  $\lambda$ cI-CTD is tethered to the  $\alpha$ -NTD.

The hybrid  $\alpha$  gene was created by replacing the gene segment encoding the  $\alpha$ -CTD with a gene segment encoding the  $\lambda$ cI-CTD. A derivative of the *lac* promoter bearing a single  $\lambda$  operator ( $O_R2$ ) in place of the CRP-binding site was created (centered 62 bps upstream of the transcription startpoint) (Figure 1B). Ordinarily,  $\lambda$ cI activates transcription  
10 when bound at a unique position centered at position -42; as expected, therefore,  $\lambda$ cI does not activate transcription from this *lac* promoter derivative.

The *lac* promoter derivative was introduced in single copy into the chromosome of *E. coli* strain MC1000 F'λcI<sup>q</sup>. Compatible vectors driving the expression of the hybrid  $\alpha$  gene and the cI gene were also introduced into this strain.  $\lambda$ cI stimulated transcription from  
15 the *lac* promoter derivative a maximum of approximately 10-fold as measured by  $\beta$ -galactosidase assays. This stimulation was observed only in the presence of the hybrid  $\alpha$  gene; in its absence  $\lambda$ cI repressed transcription slightly. Furthermore, expression of the  $\alpha$ -cI fusion protein had no significant effect on transcription from the *lac* promoter derivative in the absence of  $\lambda$ cI. Primer extension analysis confirmed that the stimulatory effect of  $\lambda$ cI  
20 reflected an increase in correctly initiated transcripts.

Our hypothesis concerning the mechanism of this activation predicts that a  $\lambda$ cI mutant unable to bind cooperatively to the DNA would be unable to activate transcription in this artificial system. To test this prediction an experiment was designed using the  $\lambda$ cI cooperativity mutant ( $\lambda$ cI-D197G) that is unable to bind cooperatively to both adjacent and  
25 separated operator sites, but is otherwise fully functional (i.e. its binding to a single operator site *in vivo* is indistinguishable from that of wild type  $\lambda$ cI). Unlike wild type  $\lambda$ cI, this mutant failed to activate transcription from the *lac* promoter derivative in the presence of the hybrid  $\alpha$  gene.

Furthermore, several  $\lambda$ cI mutants with specific but less severe cooperativity defects  
30 were also utilized in similar experiments. Substitutions N148D and R196M weaken, but do

- 45 -

not abolish, the dimer-dimer interaction responsible for cooperativity. Mutant R196M is more defective for cooperative binding than mutant N148D, and, like mutant D197G, both  $\lambda$ cI-N148D and  $\lambda$ cI-R196M behave indistinguishably from wild type  $\lambda$ cI in binding to a single operator site *in vivo*. The two mutants stimulated transcription from the *lac* promoter derivative more weakly than wild type  $\lambda$ cI, and the stronger cooperativity mutant also manifested a stronger activation defect.

The equilibrium dissociation constant for the interaction of  $\lambda$ cI dimers in solution is about  $10^{-6}$  M, and cooperative binding to DNA likely involves this same interaction. These results suggest that any protein-protein interaction of comparable strength involving a DNA-bound protein and a protein domain tethered to the  $\alpha$ -NTD would bring about transcriptional activation. The analysis of the  $\lambda$ cI cooperativity mutants indicates that the magnitude of the activation decreases as the dimer-dimer interaction is weakened. It is not known what would be the effect of increasing the strength of the dimer-dimer interaction. It will be interesting to learn how strong an interaction would result in maximal activation. It is possible that a sufficiently strong interaction might impede promoter clearance and, therefore, result in transcriptional repression rather than activation.

Our results indicate that a protein domain with no determinants for DNA-binding can mediate transcriptional activation when tethered to the  $\alpha$ -NTD simply by providing a surface that can be contacted by a DNA-bound protein. The discovery of the DNA-binding capability of the  $\alpha$ -CTD suggested that activators that interact with the  $\alpha$ -CTD might help stabilize its association with DNA at promoters that lack an UP element. In support of this idea, footprinting studies have indicated that the interaction between CRP and the  $\alpha$ -CTD at the *lac* promoter promotes the association of the  $\alpha$ -CTD with the DNA adjacent to the CRP-binding site and upstream of the promoter -35 region. This observation has prompted the proposal that other, and perhaps all, activators that interact with the  $\alpha$ -CTD function by recruiting the  $\alpha$ -CTD to the DNA. These findings, however, imply that activation can occur in the absence of this recruitment.

This new protein-protein contact alone suffices for gene activation, suggesting that a DNA-bound activator can recruit the holoenzyme to a promoter simply by touching an available target surface. These findings in *E. coli* imply that in prokaryotes, activation can

- 46 -

be elicited by a simple protein-protein contact involving a DNA-bound activator on the one hand and an available target surface within the RNA polymerase holoenzyme on the other.

λcl normally activates transcription at the λ P<sub>RM</sub> promoter using an activation patch on its N-terminal domain to contact the σ subunit of RNA polymerase. This contact 5 requires that λcl be bound just upstream of the P<sub>RM</sub> -35 region at a site centered at position -42. An experiment was designed to ask whether λcl bound at this position could use both its normal activation patch and its C-terminal domain to make simultaneous contacts with RNA polymerase in a strain expressing the α-cl fusion protein. This was found to work spectacularly well. Whereas λcl normally stimulates PRM transcription by a factor of less 10 than 10, an approximately 100-fold stimulation in a strain expressing the α-cl fusion was observed.

This finding suggests that one could use this set up to detect extremely weak protein-protein interactions. In fact, the data with the D197G mutant shows that with this assay a weak residual interaction can be detected.

15

All of the above-cited references and publications are hereby incorporated by reference.

*Equivalents*

20 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention and are covered by the following claims.

- 47 -

We claim:

1. A method for detecting interaction between a first test polypeptide and a second test polypeptide, comprising
  - i. providing an interaction trap system including a prokaryotic host cell which contains
    - 5 (a) a reporter gene operably linked to a transcriptional regulatory sequence which includes a binding site ("DBD recognition element") for a DNA-binding domain,
    - (b) a first chimeric gene which encodes a first fusion protein, said first fusion protein including a DNA-binding domain and first test polypeptide,
    - 10 (c) a second chimeric gene which encodes a second fusion protein including an activation tag activates transcription of the reporter gene when localized to the vicinity of the DBD recognition element,wherein interaction of the first fusion protein and second fusion protein in the host cell results in measurably greater expression of the reporter gene;
  - 15 ii. measuring expression of said reporter gene; and
  - iii. comparing the level of expression of said reporter gene to a level of expression in a control interaction trap system in which one of both of the first and second test polypeptides are missing from the first and second fusion proteins and resulting fusion proteins do not interact,
- 20 wherein a statistically significant increase in the level of expression is indicative of an interaction between the first and second test polypeptide portions of the fusion proteins.
2. The method of claim 1, wherein the activation tag is a polymerase interaction domain (PID) which forms active RNA polymerase complexes in the host cell
- 25 3. The method of claim 2, wherein the PID includes at least a portion of an RNA polymerase subunit.
- 30 4. The method of claim 3, wherein the PID includes at least a portion of an  $\alpha$  or  $\omega$  polymerase subunit.

- 48 -

5. The method of claim 1, wherein the host cell is selected from the group consisting of bacterial strains of *Escherichia*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia* and *Shigella*.
- 5 6. The method of claim 1, wherein the reporter gene encodes a gene product that gives rise to a detectable signal selected from the group consisting of: color, fluorescence, luminescence, cell viability relief of a cell nutritional requirement, cell growth, and drug resistance.
- 10 7. The method of claim 1, wherein the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase, luciferase,  $\beta$ -galactosidase and alkaline phosphatase.
- 15 8. The method of claim 1, wherein at least one of the first and second test polypeptides are from a nucleic acid library.
9. The method of claim 1, wherein the DNA-binding domain includes a DNA binding portion of a transcriptional regulatory protein.
- 20 10. The method of claim 1, wherein the first fusion protein also includes an oligomerization motif.
11. A kit for detecting interaction between a first test polypeptide and a second test polypeptide, the kit comprising:
  - 25 i. a first vector for encoding a first fusion protein ("bait fusion protein"), which vector comprises a first gene including:
    - (1) transcriptional and translational elements which direct expression in a prokaryotic host cell,
    - (2) a DNA sequence that encodes a DNA-binding domain and which is functionally associated with the transcriptional and translational elements of the first gene, and

- 49 -

(3) a means for inserting a DNA sequence encoding a first test polypeptide into the first vector in such a manner that the first test polypeptide is capable of being expressed in-frame as part of a bait fusion protein containing the DNA binding domain;

5       ii. a second vector for encoding a second fusion protein ("prey fusion protein"), which comprises a second gene including:

          (1) transcriptional and translational elements which direct expression in a prokaryotic host cell,

          (2) a DNA sequence that encodes a polymerase interaction domain (PID) which forms active RNA polymerase complexes in the prokaryotic host cell. the PID DNA sequence being functionally associated with the transcriptional and translational elements of the second gene, and

          (3) a means for inserting a DNA sequence encoding the second test polypeptide into the second vector in such a manner that the second test polypeptide is capable of being expressed in-frame as part of a prey fusion protein containing the polymerase interaction domain; and

15       iii. a prokaryotic host cell containing a reporter gene having a binding site ("DBD recognition element") for the DNA-binding domain, wherein the reporter gene expresses a detectable protein when a prey fusion protein interacts with a bait fusion protein bound to the DBD recognition element; the host cell being incapable of expressing any appreciable level of a protein having the function of (a) the first marker gene, (b) the second marker gene, (c) the DNA-binding domain, and (d) the polymerase interaction domain;

20       wherein binding of the first test polypeptide and the second test polypeptide in the host cell results in measurably greater expression of the reporter gene than the simultaneous presence of the DNA-binding domain and the polymerase interaction domain in the absence of an interaction between the first test polypeptide and the second test polypeptide.

25

12. The kit of claim 11, wherein the activation tag is a polymerase interaction domain (PID) which forms active RNA polymerase complexes in the host cell :

- 50 -

13. The kit of claim 12, wherein the PID includes at least a portion of an RNA polymerase subunit.
14. The kit of claim 13, wherein the PID includes at least a portion of an  $\alpha$  or  $\omega$  polymerase subunit.
15. The kit of claim 11, wherein the host cell is selected from the group consisting of bacterial strains of *Escherichia*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia* and *Shigella*.
16. The kit of claim 11, wherein the reporter gene encodes a gene product that gives rise to a detectable signal selected from the group consisting of: color, fluorescence, luminescence, cell viability relief of a cell nutritional requirement, cell growth, and drug resistance.
17. The kit of claim 11, wherein the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase, luciferase,  $\beta$ -galactosidase and alkaline phosphatase.
18. The kit of claim 11, wherein at least one of the first and second test polypeptides are from a nucleic acid library.
19. The kit of claim 11, wherein the DNA-binding domain includes a DNA binding portion of a transcriptional regulatory protein.
20. The kit of claim 11, wherein the first fusion protein also includes an oligomerization motif.
21. A method for isolating a nucleic acid encoding a polypeptide which a selected protein target, comprising
  - i. providing an interaction trap system including a vareigated population of prokaryotic host cell which each include:

**SUBSTITUTE SHEET (RULE 26)**

- 51 -

- (a) a reporter gene operably linked to a transcriptional regulatory sequence which includes a binding site ("DBD recognition element") for a DNA-binding domain,
- (b) a first chimeric gene which encodes a first fusion protein, said first fusion protein including a DNA-binding domain and first test polypeptide,
- (c) a second chimeric gene which encodes a second fusion protein including an activation tag activates transcription of the reporter gene when localized to the vicinity of the DBD recognition element,  
wherein interaction of the first fusion protein and second fusion protein in the host cell results in measurably greater expression of the reporter gene, and one of the first or second chimeric genes is present in the host cell population as a variegated population with respect to sequence encoding test polypeptides;

- ii. measuring expression of said reporter gene under conditions wherein a statistically significant increase in the level of expression of the reporter gene is indicative of an interaction between the first and second test polypeptide portions of the fusion proteins; and
- iii. selecting cells from the host cell population on the basis of the level of expression of said reporter gene.

20 22. The method of claim 21, wherein the activation tag is a polymerase interaction domain (PID) which forms active RNA polymerase complexes in the host cell

25 23. The method of claim 22, wherein the PID includes at least a portion of an RNA polymerase subunit.

24. The method of claim 23, wherein the PID includes at least a portion of an  $\alpha$  or  $\omega$  polymerase subunit.

30 25. The method of claim 21, wherein the host cell is selected from the group consisting of bacterial strains of *Escherichia*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia* and *Shigella*.

- 52 -

26. The method of claim 21, wherein the reporter gene encodes a gene product that gives rise to a detectable signal selected from the group consisting of: color, fluorescence, luminescence, cell viability relief of a cell nutritional requirement, cell growth, and drug 5 resistance.

27. The method of claim 21, wherein the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase, luciferase,  $\beta$ -galactosidase and alkaline phosphatase.

10 28. The method of claim 21, wherein the DNA-binding domain includes a DNA binding portion of a transcriptional regulatory protein.

29. The method of claim 21, wherein the first fusion protein also includes an 15 oligomerization motif.

1/8

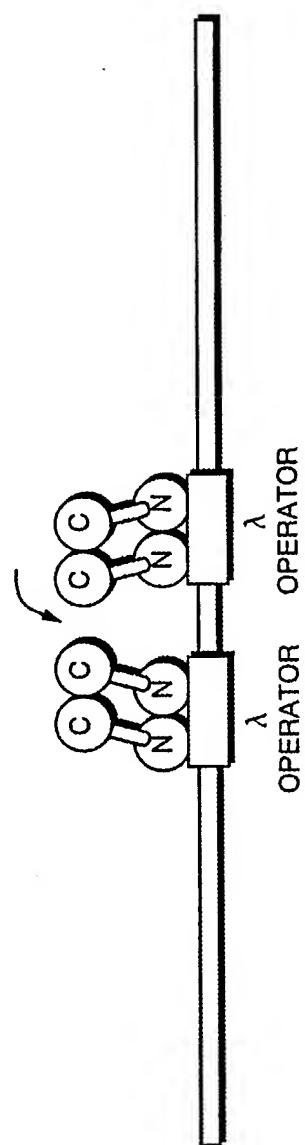


FIG. 1A

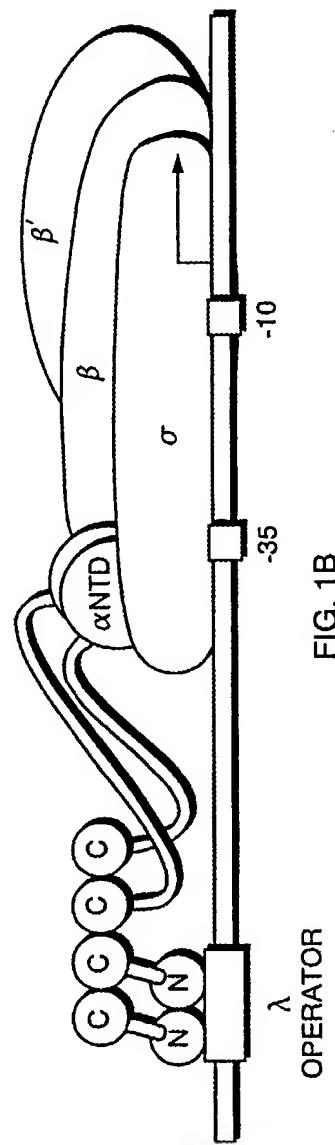


FIG. 1B

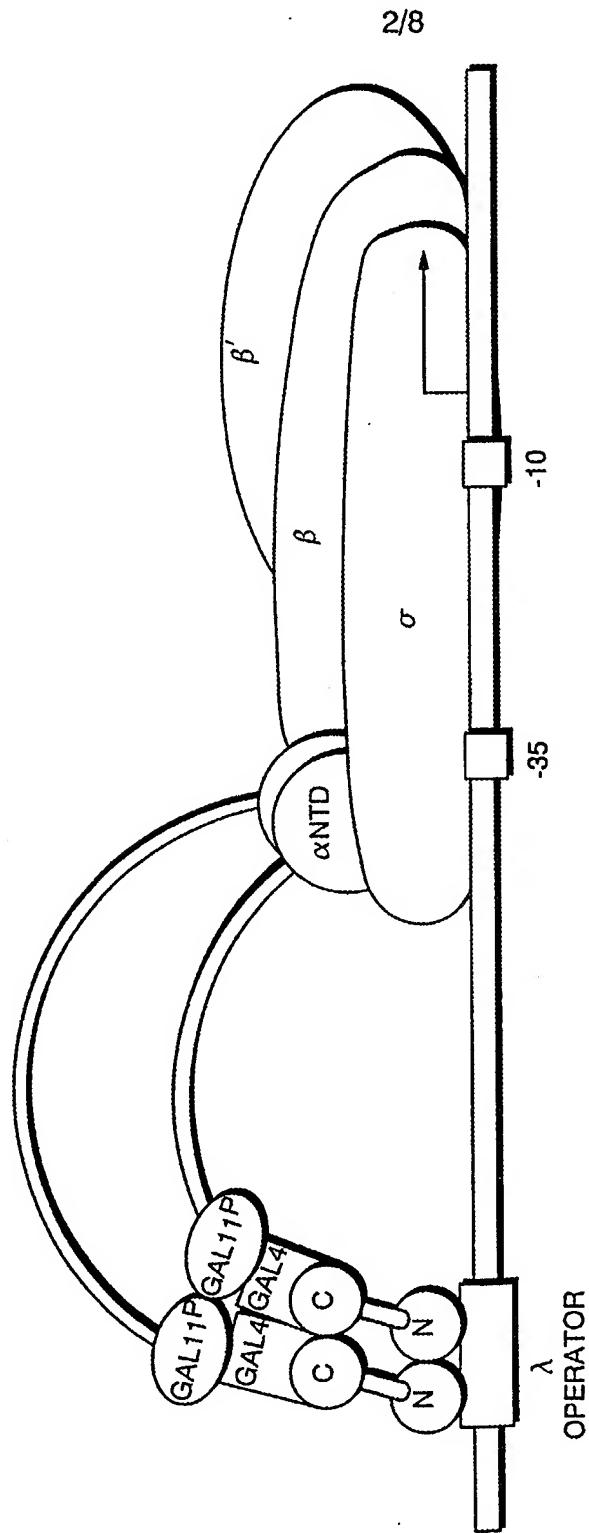
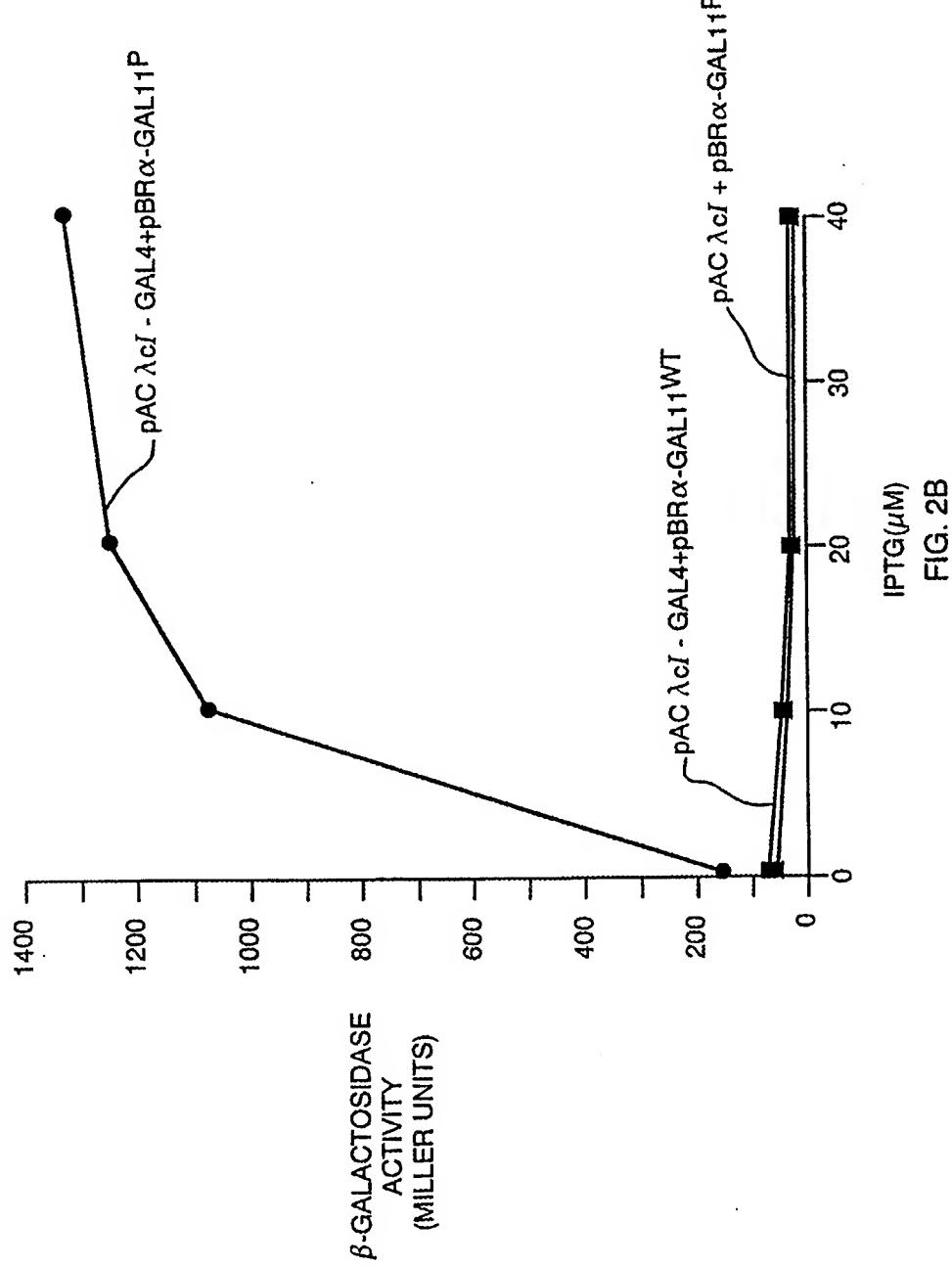


FIG. 2A

3/8



SUBSTITUTE SHEET (RULE 26)

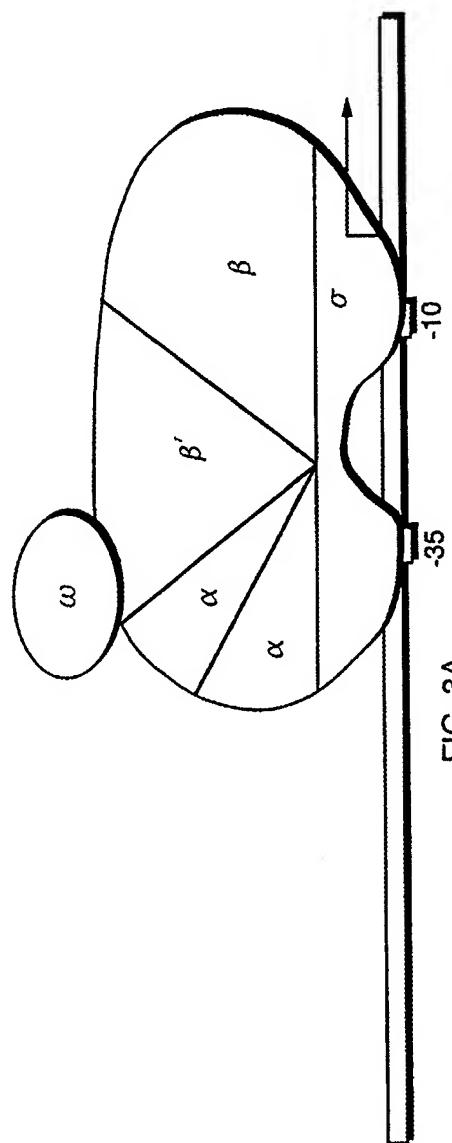


FIG. 3A

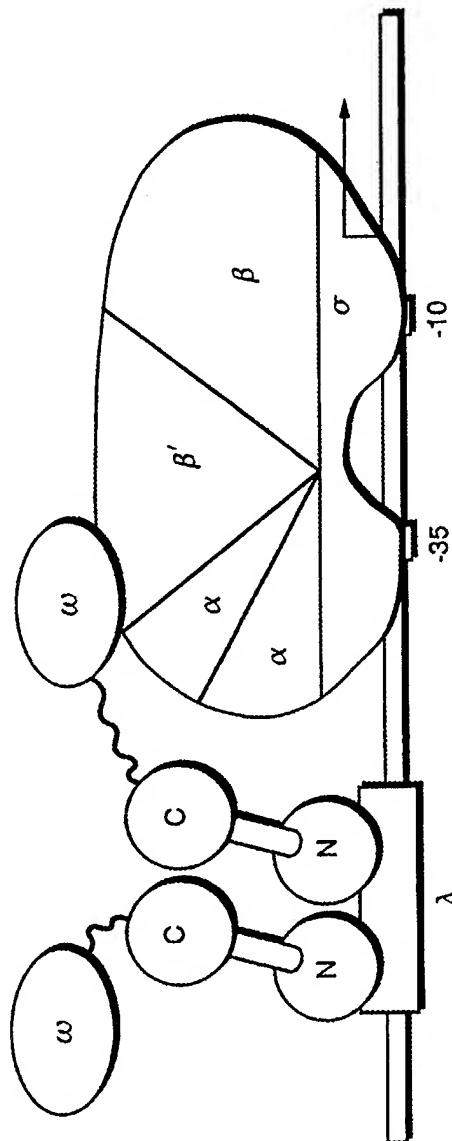
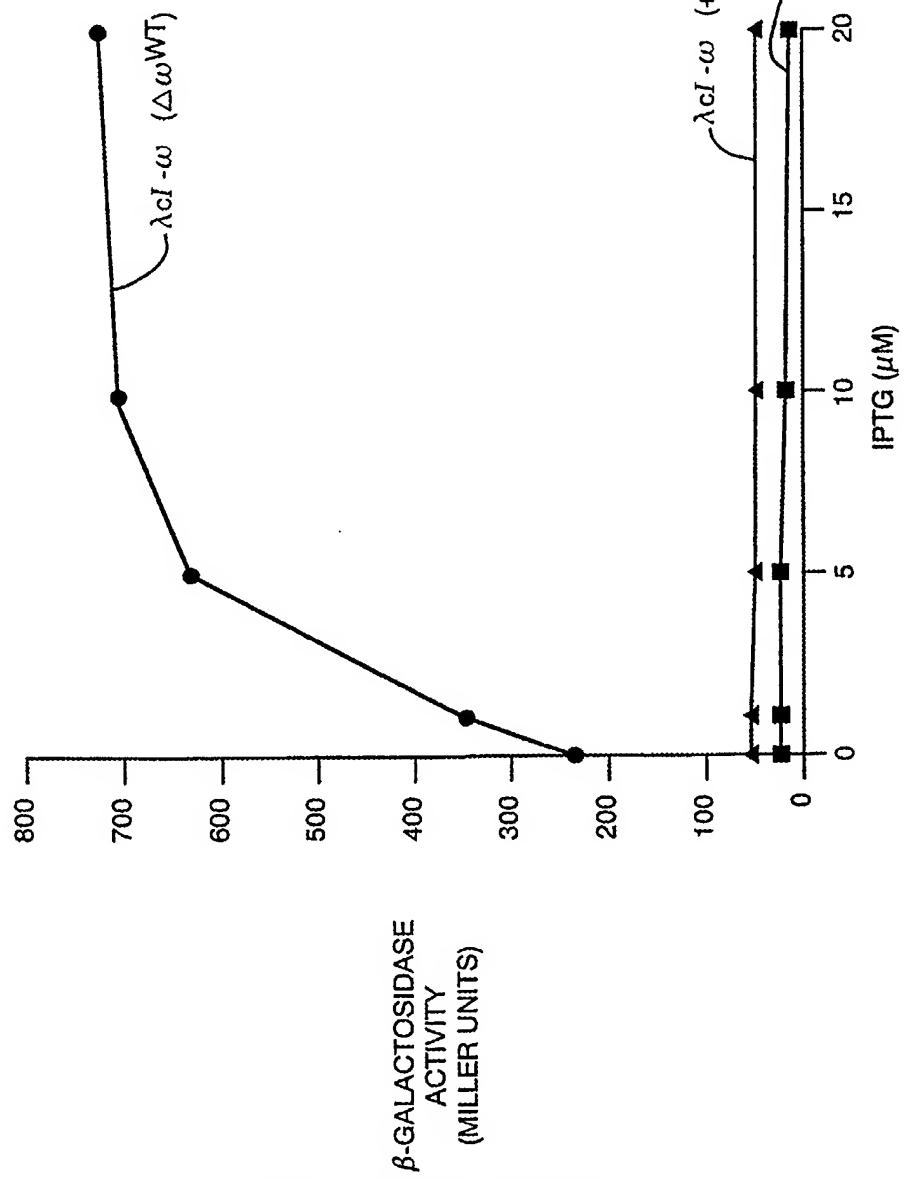


FIG. 3B

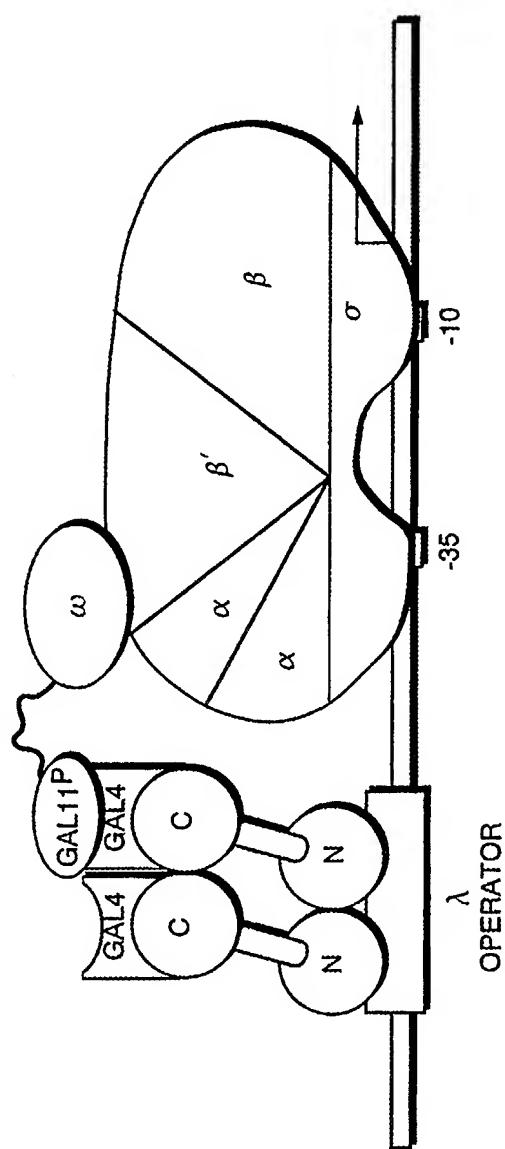
5/8



SUBSTITUTE SHEET (RULE 26)

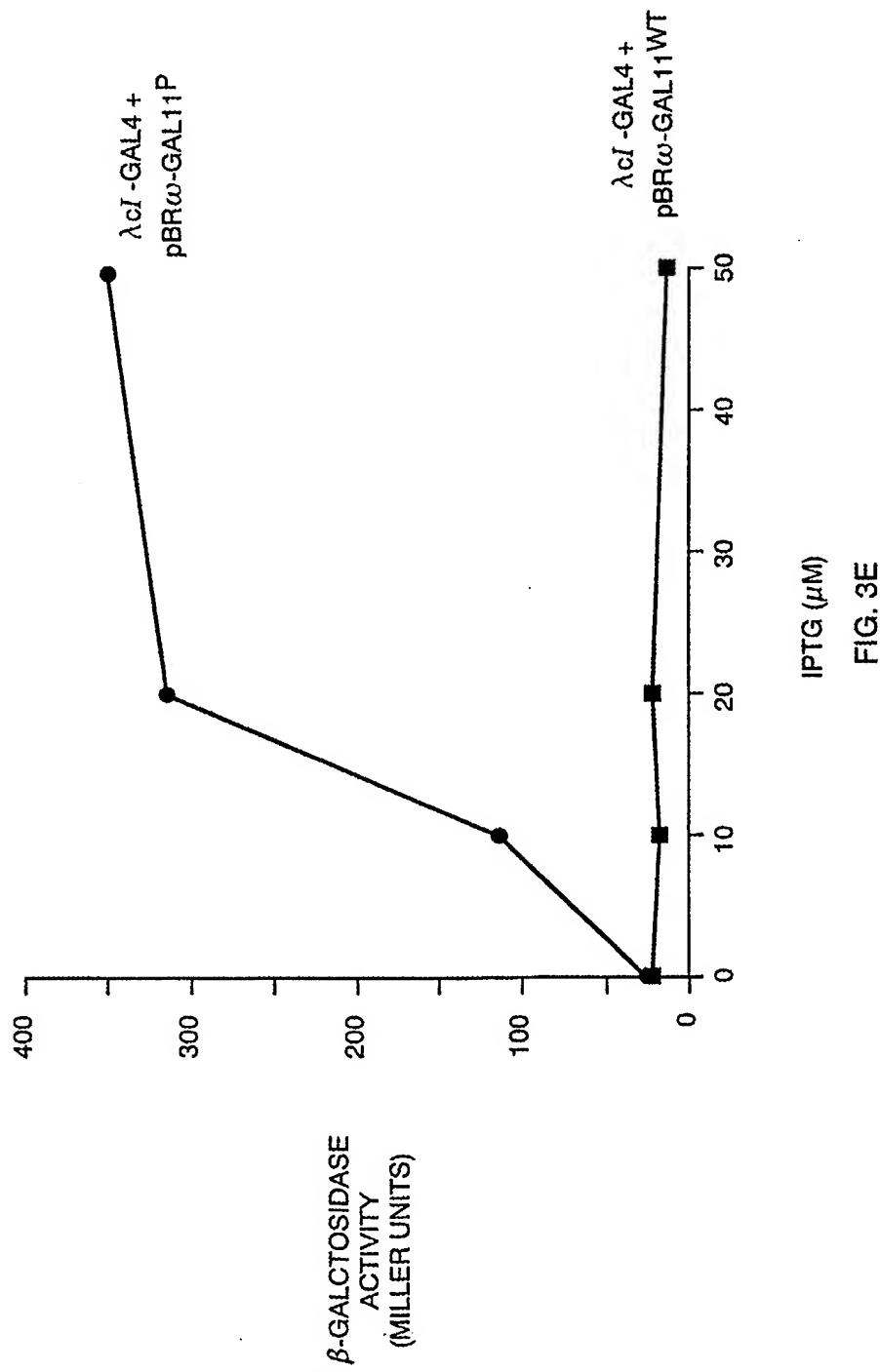
FIG. 3C

6/8



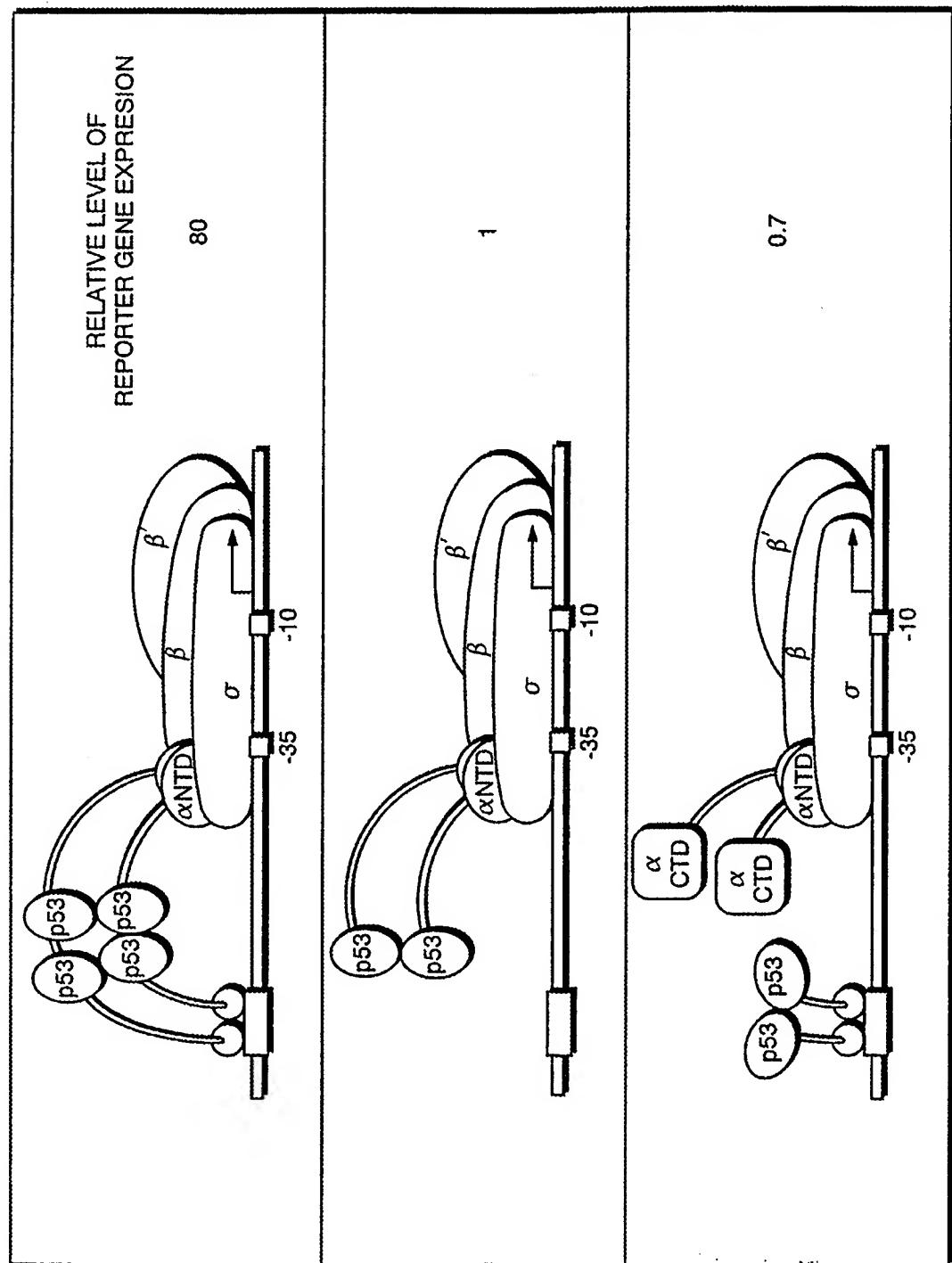
SUBSTITUTE SHEET (RULE 26)

7/8



SUBSTITUTE SHEET (RULE 26)

8/8



SUBSTITUTE SHEET (RULE 26)

FIG. 4

# INTERNATIONAL SEARCH REPORT

Internal Application No  
PCT/US 97/14860

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 C12N15/10 C12Q1/68 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BARBERIS ET AL.: "Contact with a component of the polymerase II holoenzyme suffices for gene activation" CELL, vol. 81, 1995, pages 359-368, XP002050706 See page 366, right column see the whole document	1-29
Y	CHATTERJEE AND STRUHL: "Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain" NATURE, vol. 374, 1995, pages 820-822, XP002050707 see Fig.2 see the whole document	1-29
-/-		

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance  
"E" earlier document but published on or after the International filing date  
"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
"O" document referring to an oral disclosure, use, exhibition or other means  
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
"S" document member of the same patent family

1

Date of the actual completion of the international search

19 December 1997

Date of mailing of the international search report

16/01/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

## INTERNATIONAL SEARCH REPORT

Internal Application No  
PCT/US 97/14860

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 283 173 A (FIELDS STANLEY ET AL) 1 February 1994 see the whole document -----	1-29
A	WO 94 10300 A (GEN HOSPITAL CORP) 11 May 1994 see page 13 - page 19 -----	1-29
A	ISHIHAMA: "Protein-protein communication within the transcription apparatus" JOURNAL OF BACTERIOLOGY, vol. 175, no. 9, 1993, pages 2483-2489, XP002050862 see the whole document -----	1-29
P,X	DOVE ET AL.: "Activation of prokaryotic transcription through arbitrary protein-protein contacts" NATURE, vol. 386, 1997, pages 627-630, XP002050708 see the whole document -----	1-29

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

Intern. Appl. No.

PCT/US 97/14860

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5283173 A	01-02-94	US 5468614 A US 5667973 A	21-11-95 16-09-97
WO 9410300 A	11-05-94	EP 0672131 A JP 8506480 T US 5580736 A	20-09-95 16-07-96 03-12-96